Content-based Information Retrieval and Anonymisation in Data and Multimedia Streams

Athanasios Zigomitros

Thesis submitted for the degree of

Doctor of Philosophy

University of Piraeus 2018

This page is intentionally left blank.

Content-based Information Retrieval & Anonymisation in Data & Multimedia Streams

Thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

in the Department of Informatics
of the School of Information and Communication Technologies
at the University of Piraeus

By

Athanasios Zigomitros

Supervising Committee

Constantinos Patsakis
University of Piraeus

Georgios Tsihrintzis
University of Piraeus

Dimitrios Apostolou
University of Piraeus

Approved by

Date: October, 8th 2018.

……………………
Constantinos Patsakis
Assistant Professor
University of Piraeus

……………………
Georgios Tsihrintzis
Professor
University of Piraeus

……………………
Dimitrios Apostolou
Associate Professor
University of Piraeus

……………………
Yannis Theodoridis
Professor
University of Piraeus

……………………
Vasilios Katos
Professor
Bournemouth University

……………………
Georgios Kambourakis
Associate Professor
Aegean University

……………………
Emmanouil Magkos
Associate Professor
Ionian University

This page is intentionally left blank.

This thesis is dedicated to
my parents, my brothers and my wife
for their continuous support.

This page is intentionally left blank.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Constantinos Patsakis for the continuous support of my PhD study, for his endless patience and motivation. It is hard to even imagine having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Georgios Tsihrintzis, and Associate Prof. Dimitris Apostolou for their insightful comments and encouragement.

My sincere thanks also go to Prof. Dimitris Despotis who gave me the opportunity to start this journey and for supporting me when I decided to change course on my research.

I thank my friends at the University of Piraeus, Achilleas Papageorgiou, Panagiotis Tampakis, Dr Dimitris Sotiros, Angeliki Mikeli, Dr Gregory Koronakos, Charis Katsavos and Despina Kopanaki. Without their precious support, it would not be possible to conduct this research.

From the Institute for the Management of Information Systems at Research Center "Athena", I would like to thank Dr Olga Gkountouna and Dr Manolis Terrovitis for our co-operation and the new research interest which they helped me explore.

For their valuable contribution to my research, I would also like to thank Prof. Agusti Solanas, Dr Fran Casino and Ioannis Chrisovergis.

I would like to thank my Professors in TEI of Patras Emmanouil Z. Psarakis, Apostolos Rafailidis and Mitropoulos Ioannis who inspired me to follow the difficult path of research.

Last but not the least, I would like to thank my family: my parents, my brothers and my wife for supporting me spiritually throughout writing this thesis and my life in general.

And thank you, reader of this thesis.

September 2018,

Athanasios Zigomitros

This page is intentionally left blank.

Abstract

Personal data is any information that can be used to identify a person. These data can take different forms such as a field in a database, a unique number, a photograph or a network packet. Personal data about individuals is collected and manage from organizations in private and public sector. Businesses and the scientific community are both hungry for data. With the advanced algorithms of data mining original knowledge can be revealed from this data. Therefore, these data cannot be disseminated carelessly because the danger of breaching individuals' privacy is always present. A scientific area of data anonymisation was emerged to protect the privacy and several methods have been proposed to guarantee data privacy in published datasets. There is a trade-off to apply anonymisation algorithms. The anonymisation process should balance between the protection of the privacy of the individuals and the usefulness of the released dataset. These anonymisation technics are analyzed and their strong and weak points are highlighted. The contribution of this thesis on data publishing scentific field is twofold. First, the introduction of a new attack on anonymised data, called *inference of $\mathcal{QI}s$ attack*, which shows that an automated anonymisation solution, especially for medical records, is difficult without taking into account the semantics of the data and without consulting experts in this field. Second, the development of an algorithm which implements the k^m-anonymisation by taking into account the properties of continuous attributes and without giving a generalisation hierarchy. We conduct experiments which show that our algorithm preserves more information in the published dataset in comparison to other anonymisation algorithms that use generalisation hierarchy trees.

Multimedia is another type of personal data that also examined in this research. From multimedia that is shared on Online Social Networks derives multiple privacy risks. An analytic survey of these risks is presented and a solution based on digital watermarking has been proposed towards to elimination of many of these risks.

Smartphones increasing compute capabilities are paired with sensors, such as GPS, offering new opportunities to develop mobile applications with new potentials. To demonstrate how the privacy of a user can be breached we focused on a privacy-sensitive domain of apps, the dating apps. The research is based on the transmitted network packets and the results are worrying.

Keywords: Data Anonymisation, Privacy-Preserving Data Publishing, k-anonymity, Inference of QIs Attack, k^m-anonymity, Multimedia Privacy, Watermarking , Mobile Application Security

Abstract

Προσωπικά δεδομένα είναι οποιαδήποτε πληροφορία μπορεί να χρησιμοποιηθεί για την αναγνώριση ενός ατόμου. Αυτά τα δεδομένα μπορούν να λάβουν διάφορες μορφές, όπως ένα πεδίο σε μια βάση δεδομένων, έναν μοναδικό αριθμό, μια φωτογραφία ή ένα πακέτο δικτύου. Τα προσωπικά δεδομένα σχετικά με τα άτομα συλλέγονται και διαχειρίζονται από οργανισμούς στον ιδιωτικό και δημόσιο τομέα. Οι επιχειρήσεις και η επιστημονική κοινότητα έχουν μια ακόρεστη δίψα για δεδομένα. Με τους προηγμένους αλγόριθμους εξόρυξης δεδομένων νέα γνώση μπορεί να αποκαλυφθεί από αυτά τα δεδομένα. Πρέπει να σημειωθεί ότι αυτά τα δεδομένα δεν μπορούν να δημοσιευθούν χωρίς την δέουσα προσοχή αφού υπάρχει πάντα ο κίνδυνος παραβίασης της ιδιωτικής ζωής των ατόμων. Ένας επιστημονικός τομέας αναδείχθηκε, αυτός της ανωνυμοποίησης δεδομένων, για να καλύψει την ανάγκη της προστασίας της ιδιωτικότητας σε δεδομένα που δημοσιεύονται. Η διαδικασία της ανωνυμοποίησης θα πρέπει να εξισορροπήσει την προστασία της ιδιωτικότητας των ατόμων με τη χρησιμότητα του δημοσιευμένου συνόλου δεδομένων. Αυτές οι τεχνικές ανωνυμοποίησης αναλύονται και παρουσιάζονται τα πλεονεκτήματα και μειονεκτήματα τους. Η συμβολή αυτής της διατριβής στο επιστημονικό πεδίο της δημοσίευσης δεδομένων είναι διττή. Πρώτον, την εισαγωγή μιας νέας επίθεσης σε ανώνυμοποιημένα δεδομένα, που ονομάζεται επίθεση συμπερασμού των οιονεί αναγνωριστικών, *inference of $\mathcal{QI}s$ attack*, που δείχνει ότι μια αυτοματοποιημένη λύση ανωνυμοποίησης, ειδικά για ιατρικά δεδομένα, είναι δύσκολο να επιτευχθεί χωρίς να ληφθεί υπόψη η σημασιολογία των δεδομένων και χωρίς την συμβολή των εμπειρογνωμόνων του τομέα. Δεύτερον, η ανάπτυξη ενός νέου αλγορίθμου που υλοποιεί k^m-anonymisation λαμβάνοντας υπόψη τις ιδιότητες των συνεχών χαρακτηριστικών και χωρίς να προαπαιτεί μια ιεραρχία γενίκευσης. Μετά απο διεξαγωγή πειραμάτων φαίνεται ότι ο νεος αλγόριθμός διατηρεί περισσότερες πληροφορίες στο δημοσιευμένο σύνολο δεδομένων σε σύγκριση με άλλους αλγόριθμους ανωνυμοποίησης που χρησιμοποιούν ιεραρχίες γενίκευσης.

Τα πολυμέσα είναι ένας άλλος τύπος προσωπικών δεδομένων που επίσης εξετάστηκε σε αυτή την έρευνα. Συγκεκριμένα απο τα πολυμέσα που διαμοιράζονται στα Online Κοινωνικά Δίκτυα προκύπτουν πολλοί κίνδυνοι ιδιωτικότητας. Παρουσιάζεται μια αναλυτική έρευνα για τους κινδύνους αυτούς και προτείνεται μια λύση με βάση την ψηφιακή υδατογράφηση για την εξάλειψη πολλών από αυτούς τους κινδύνους.

Τα έξυπνα τηλέφωνα με την συνεχως αυξανομενη υπολογιστικη δυναμη συνδυάζονται με αισθητήρες, όπως το GPS, προσφέροντας νέες ευκαιρίες για ανάπτυξη κινητών εφαρμογών με νέες δυνατότητες. Για να δείξουμε πώς μπορεί να παραβιαστεί το απόρρητο ενός χρήστη, επικεντρώσαμε την προσοχή μας σε έναν τομέα ευαίσθητο στην ιδιωτικότητα των εφαρμογών, των εφαρμογών που χρονολογούνται. Η έρευνα βασίζεται στα μεταδιδόμενα πακέτα δικτύου και τα αποτελέσματα είναι ανησυχητικά.

Λεξεις κλειδια: Ανωνυμοποίηση Δεδομένων, k-ανωνυμία, Επίθεση Συμπερασμού Οιονεί Αναγνωριστικών, k^m-ανωνυμία , Ιδιωτικότητα Πολυμέσων, Υδατογράφηση, Ασφάλεια Εφαρμογών Κινητών Τηλεφώνων

Contents

List of Figures

This page is intentionally left blank.

Part I

Prelude

This page is intentionally left blank.

Chapter 1

Introduction

This thesis examines privacy implications of Content-based Information Retrieval. Content-based Information Retrieval has a different meaning and must not be confused with CBIR (Content-Based Image Retrieval). It refers to the content itself that can lead to privacy breaches. For example, this can be a data string that reveals sensitive information about a person or a small detail on an image that can pinpoint his exact location or even a network packet that carries more information than it shows on the screen. Under the new General Data Protection Regulation (GDPR) [1] the definition of personal data is the following:

> "Personal data" means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

This definition does not leave any space for false interpretations of what personal data are. Any information that can be used on its own or combined with other information to identify, contact, or locate an individual. Hence, in the EU, any location information or any device ID which uniquely bind a place or a device respectively to an individual fall under the personal data category.

The GDPR is a legislation to protect the personal information from misuse and to allow citizens have better control of their data e.g. by allowing them the Right to be

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

Forgotten [198]. GDPR concerns the E.U. citizens, but it is reshaping the privacy domain worldwide since any company doing business with EU citizens has to comply. Due to its scope, GDPR has many implications to numerous sectors, spanning from all private to all public sector applications, and disrupting common application and implementation workflow [199].

Due to the ever increasing amount of information that is produced in daily basis [110], a huge amount of data is hiding knowledge waiting to be discovered. Healthcare organizations are a good example of this kind of data which can be utilized for the better treatment of the patient, medical diagnosis, and other health-related discoveries. Apparently, this kind of data should not be published or shared without proper anonymisation. The current state of art in Privacy-Preserving Data Publishing offers a lot of different approaches with each one having its advantages and disadvantages which are going to be analysed in this thesis.

Another alarming example of misuse of personal data is the recent Facebook–Cambridge Analytica data scandal [2]. Cambridge Analytica by developing a Facebook personality quiz app, which called "thisisyourdigitallife", collected data from 270.000 app users and all their friends in their friend list. The affected profiles estimated to 50 million users. The users of the app had given implied consent to the collection of their personal data but their friends had almost certainly not. As is apparent Online Social Network holding a lot of personal data and therefore the security and privacy of them is examined by many researchers [98, 290, 88, 118, 294, 25, 80, 207, 129]. This thesis examines the security and privacy risks which arise from the sharing of multimedia on online social networks as the previous studies did not gave the proper attention on multimedia.

1.1 Thesis structure

This thesis starts with an analytic introduction to anonymisation of tabular data. Chapter 2 provides the fundamentals of privacy-preserving data publication. In Chapter 3 demonstrates how the properties of the data or their semantics could allow an attacker to invade the privacy of the record owners in an anonymised dataset. A new modified algorithm which offers k^m-anonymity without the need of generalisation hierarchy is introduced in 4.

[2] https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Multimedia is a key component in the majority of Online Social Networks. In Chapter 5 we discuss the privacy breaches that could be made from uploaded multimedia data in Online Social Networks. All possible attacks are recorded and analysed and subsequently in Chapter 6 a watermarking scheme is proposed to solve some of the aforementioned problems. This scheme can be used inside one social network or in a distributed way on multiple social networks to enforce privacy settings on multimedia across collaborating Online Social Networks.

Chapter 7 examines the security and privacy of mobile applications. To demonstrate a real life scenario on how the privacy is violated in mobile applications this thesis analyses the results of a thorough study of mobile dating applications, which due to their nature can be considered as a social network where users exchange very sensitive personal data.

Chapter 8 presents the open research questions which need further investigation and finally, Chapter 9 contains the conclusions of the dissertation.

The thesis ends with the full bibliography of the publications and books used as in this thesis.

1.2 Research Projects

During my thesis, I participated in several EU and national funded R&D projects which have significantly affected the conducted research. These projects are the following:

- OPERANDO Online Privacy Enforcement, Rights Assurance & Optimization operando.eu. Supported by the European Commission under the Horizon 2020 Programme (H2020), as part of the OPERANDO project (GA no. 653704).

- IDEA-C http://idea-c.weebly.com/ Co-funded by the Europe for Citizens Program of the European Union.

- GEOSTREAM: Exploiting User-Generated Geospatial Content Streams http://geocontentstream.eu. GEOSTREAM is partially supported by the FP7 - Research for SMEs programme of the European Commission under contract number FP7-SME-2012-315631.

- THALES: Methodological expansions of Data Envelopment Analysis and application in the evaluation of Greek Universities - Operational Program ”Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF), 2012-2015

1.3 List of publications

Part of the findings during my research for this thesis have been published in peer reviewed journals and conferences. More precisely, the following publications were made in JCR indexed journals:

- Constantinos Patsakis, Athanasios Zigomitros, Achilleas Papageorgiou, and Agusti Solanas. Privacy and security for multimedia content shared on osns: Issues and countermeasures. *Comput. J.*, 58(4):518–535, 2015

- Constantinos Patsakis, Athanasios Zigomitros, Achilleas Papageorgiou, and Edgar Galván López. Distributing privacy policies over multimedia content across multiple online social networks. *Computer Networks*, 75:531–543, 2014

Moreover, the following publications have been made in peer reviewed international conferences:

- Athanasios Zigomitros, Achilleas Papageorgiou, and Constantinos Patsakis. A practical k-anonymous recommender system. In *Information, Intelligence, Systems & Applications (IISA), 2016 7th International Conference on*, pages 1–4. IEEE, 2016

- A. Papageorgiou, A. Zigomitros, and C. Patsakis. Personalising and crowdsourcing stress management in urban environments via s-health. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–4, July 2015

- Constantinos Patsakis, Athanasios Zigomitros, and Agusti Solanas. Privacy-aware genome mining: Server-assisted protocols for private set intersection and pattern matching. In Caetano Traina Jr., Pedro Pereira Rodrigues, Bridget Kane, Paulo Mazzoncini de Azevedo Marques, and Agma Juci Machado Traina, editors, *28th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2015, Sao Carlos, Brazil, June 22-25, 2015*, pages 276–279. IEEE, 2015

- Constantinos Patsakis, Athanasios Zigomitros, and Agusti Solanas. Analysis of privacy and security exposure in mobile dating applications. In Selma Boumerdassi, Samia Bouzefrane, and Éric Renault, editors, *Mobile, Secure, and Programmable Networking - First International Conference, MSPN 2015, Paris, France, June 15-17, 2015, Selected Papers*, volume 9395 of *Lecture Notes in Computer Science*, pages 151–162. Springer, 2015

- Constantinos Patsakis, Michael Clear, Paul Laird, Athanasios Zigomitros, and Mélanie Bouroche. Privacy-aware large-scale virological and epidemiological data monitoring. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014* [4], pages 78–81

- Athanasios Zigomitros, Agusti Solanas, and Constantinos Patsakis. The role of inference in the anonymization of medical records. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014* [4], pages 88–93

- Olga Gkountouna, Sotiris Angeli, Athanasios Zigomitros, Manolis Terrovitis, and Yannis Vassiliou. k^m-anonymity for continuous data using dynamic hierarchies. In Josep Domingo-Ferrer, editor, *Privacy in Statistical Databases*, volume 8744 of *Lecture Notes in Computer Science*, pages 156–169. Springer International Publishing, 2014

- Athanasios Zigomitros and Constantinos Patsakis. Storing metadata as QR codes in multimedia streams. In Emmanouel Garoufallou and Jane Greenberg, editors, *Metadata and Semantics Research - 7th Research Conference, MTSR 2013, Thessaloniki, Greece, November 19-22, 2013. Proceedings*, volume 390 of *Communications in Computer and Information Science*, pages 152–162. Springer, 2013

- Athanasios Zigomitros, Achilleas Papageorgiou, and Constantinos Patsakis. Social network content management through watermarking. In Geyong Min, Yulei Wu, Lei (Chris) Liu, Xiaolong Jin, Stephen A. Jarvis, and Ahmed Yassin Al-Dubai, editors, *11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2012, Liverpool, United Kingdom, June 25-27, 2012*, pages 1381–1386. IEEE Computer Society, 2012

- Athanasios Zigomitros and Constantinos Patsakis. Cross format embedding of metadata in images using qr codes. In *Intelligent Interactive Multimedia Systems and Services*, volume 11 of *Smart Innovation, Systems and Technologies*, pages 113–121. Springer, 2011

This page is intentionally left blank.

Part II

Data Anonymisation

This page is intentionally left blank.

Chapter 2

Introduction to Data Anonymisation

Knowledge discovery is a key element to innovation. Researchers analyse data and come up with new discoveries. The patterns of human behaviour are explored by companies in order to deliver better products and services to their customers. Nowdays, in the age of Big Data , the huge amount of knowledge is hidden in the electronic traces of human activity. Aside from the positive side, many concerns have been raised regarding the privacy of people, as the combination of available information can recover sensitive attributes. An adversary may abuse this knowledge, attempting to acquire sensitive data about people, and information that would not be accessible to him under other conditions, consequently invading the privacy of people [17]. A naïve way to deal with this problem is to erase the fields which are explicit identifiers to a person, for example Name, SSN etc. Nevertheless, this method has been proven to be inefficient. A person of interest for the adversary can uniquely identified even when the explicit identifiers have been removed as shown in the successful attack of Sweeney [238]. In this attack the first dataset contained the voters registration list while the second one was a dataset of patients which was provided by the Group Insurance Commission (GIC). Sweeney using three common fields Sex, Zip code and the date of birth, as depicted in Figure 2.1, linked those to two datasets. Based on the 1990 census data of US population [236] Sweeney showed that the 87% of the population can be uniquely identified through these three fields. A later work [95] on the 2000 census data decreased the percentage of the US population that is uniquely identifiable by the same fields to 63%. More studies [128, 65] conclude to similar results for other countries.

From a legal point of view, the personally identifiable information has several interpretations. For instance, the California Senate Bill 1386, defines personal identifiable information as Social Security numbers (SSN), driver's license numbers, financial accounts, excluding licence plates, credit card number, email addresses or telephone numbers. On the other hand, in 1995 the European Union gave a broader meaning to personally identifiable information as:

> "Any information relating to an identified or identifiable natural person...; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.
> ...account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person." [72]

The Data Protection Directive is superseded by the General Data Protection Regulation (GDPR), which was adopted by the European Parliament and European Council in April 2016 and is enforceable since May 2018. The definition of GDPR for personal data already mentioned in Introduction 1.

Nonetheless, as Narayanan and Shmatikov [168] point out:

> "Any information that distinguishes one person from another can be used for re-identifying anonymous data."

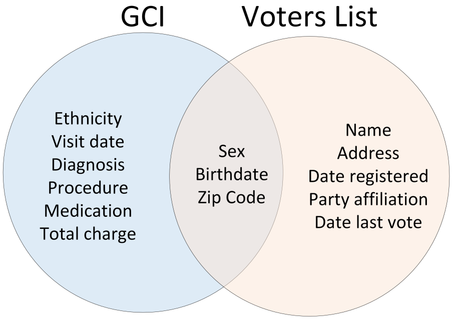


Figure 2.1: Sweeney's Example of Re-identification

The publication of any type of raw data without taking into account the implications of possible data correlations can lead to serious privacy violations from data leakages. Therefore, data publishers should apply efficient data anonymisation methods with proper guarantees in order to protect the privacy of the individuals. In this regard, this chapter focuses on the anonymisation of relational data since relational data are the most common type of published datasets.

2.1 Related work

A related field is Privacy-Preserving Data Mining (PPDM) which focus on performing the data mining tasks in a privacy-preserving way while Privacy-Preserving Data Publishing (PPDP) cares for the usefulness of the data even if they are examined on record level and before undergo any analysis. Therefore in PPDP the *truthfulness on record level* is usually a requirement, which is not the case in PPDM. The data publisher usually does not have the expertise to apply the optimal data minings algorithms and in some case the selection of the best algorithm depends on the application scenario. Moreover, when the data mining process is known beforehand then the data publisher can do that "in-house" and publish only the results. PPDP does not make any assumption regarding the type of analysis that the anonymised dataset will undergo.

To simply put the main difference between PPDM and PPDP is where the query is performed. In PPDM, the query is performed within a controlled environment, therefore, it is easier to have the necessary "watchdogs" to control the information flow and preserve the privacy of individuals. However, in the case of PPDP, the released dataset is available to the adversary to perform any query he wishes to. Therefore, PPDP needs to provide methods prior to release, whereas PPDM can be more dynamic and introduce control on the fly. For more on PPDM, the interested reader may refer to [254, 8, 155, 282].

2.2 Organisation of this Chapter

The rest of this Chapter is organised as follows. The next section introduces the reader to the basic concepts of PPDP, defining the actors and their roles in the procedure as well as the publication process. Section 2.4 presents the core data transformation methods, the building blocks which are used to anonymise datasets. The use of these methods implies some information loss, as in the processed data we need to remove the identifiable information of individuals. To quantify the amount of lost information, we detail in Section 2.5 several well-known information loss metrics. Section 2.6 presents the state of the art in anonymisation attacks and countermeasures. The goal of this section is twofold. First it shows with practical examples the different attack scenarios that an adversary would try to exploit in order to de-anonymise, even partially, the published dataset. The goal here is to highlight the possible data leakages and what is the exposure of individuals in each case. Second, based on these attacks, we present, where possible, efficient countermeasures. Nonetheless, the attack surface is bigger, therefore, in Section 2.7 we present other attacks which target the procedure per se, focusing also on the case of continuous data publishing.

2.3 Fundamentals of Privacy-Preserving Data Publishing

The goal of Privacy-Preserving Data Publishing (PPDP), as the name indicates, is to publish useful information while respecting the privacy of the people in the dataset. Therefore, the transformation of the data is essential so that an adversary can not link records of the dataset with specific individuals. The risk of privacy breaches is high when the data are carelessly disseminate. Table 2.1 is an extended version of the attacks published in the paper of El Emam et al. [67] summarising the known attacks, made mainly by researchers, on real datasets and their efficiency. It is apparent that the low number of these attacks is hard to justify the research in the field. Nevertheless, as Sweeney points out [239], such works face many problems in publication procedure, in the fear of legal measures. Moreover, one can assume that these attacks are only the tip of the iceberg since is safe to assume that majority of the attacks made by adversaries have never surfaced. Contrary to popular belief, the attack does not have to be launched from a "hacker". For instance, an attack can be launched from an ad service provider to re-identify some individuals and provide them with targeted advertisements, without the victim being aware of the attack or trace the source of the leakage. Another more malicious example is the re-identification of some individuals from an adversary which can cause attacks on individuals, e.g. ransom, who might not disclose this as they cannot trace back the origin of data leakage. The attack on Ashley Madison[1], even if it does not fall in the category of the attack on published data, clearly demonstrate how malicious people can use disclosed data to extort individuals. Nonetheless, the attacks of Table 2.1 clearly shows that there is a trade-off between the privacy and the introduced distortion on the dataset. PPDP tries to balance between privacy and utility of the anonymised dataset.

Now follows the basic concepts and definitions of Privacy-Preserving Data Publishing and several attacks which highlight the necessity for the introduction of special anonymisation algorithms.

[1] http://www.reuters.com/article/us-ashleymadison-cybersecurity-idUSKCN0QN2BN20150819

Reference	Year	Number of individuals re-identified	Proper de-identification of attacked data ?	Re-identification verified ?
German employment statistics (register data) and the German Life History Study (survey data) [21]	2001	29 of 273	Factually anonymous	Yes (records containing insurance numbers only)
Comparing Chicago homicide dataset with records in the Social Security Death Index [177]	2001	75% Of 11,000	Direct identifiers removed	No
Group Insurance Commission patient-specific data with nearly one hundred attributes and voter registration list for Cambridge Massachusetts[238]	2002	1 of 135,000	Removal of names and addresses	Yes
[69]	2003	219 unique matches, 112 with 2 possibilities, 8 confirmed	Yes	Verified matches, but not identities
Web search queries collected by AOL [17]	2006	1 of 657,000	No	Yes (with individual)
Maps of Boston with the addresses of patients plotted as individual dots or symbols that was in papers of medical journals [29]	2006	79% of 550	No	Verified (with original data set)
A set of movie ratings and a set of movie mentions from MovieLens[81]	2006	Of 133 users, 60% of those who mention at least 8 movies	Direct identifiers removed	No
[233]	2006	18 of 20	Only type of cancer, zip code and date of diagnosis included in request	Yes (verified by the Department of Health)
Identification of genomic risk for specific family relations [156]	2006	70%	No	Yes
The network of friendship links on the blogging site LiveJournal. [14]	2007	2,400 of 4.4 million	Identifying information removed	Verified using original data
[77]	2007	1	Direct Identifiers removed & possibly other unknown de-id methods used	Yes
Anonymous movie ratings of 500000 subscribers of Netflix and 50 IMDB users [166]	2008	2 of 50	Direct identifiers removed & maybe perturbation	No
[68]	2009	1 of 3,510	Direct identifiers removed	Yes
Users have accounts on both Twitter and Flickr, re-identified in the anonymous Twitter graph [167]	2009	30.8% of 150 pairs of nodes	Identifying information removed	Verified using ground-truth mapping of the 2 networks
[130, 131]	2010	2 of 15,000	Yes - HIPAA Safe Harbor	Yes
A complete dump of historical trip and fare logs from NYC taxis[a]	2014	100%	Medallion and licence numbers obfuscated	Verified using rainbow tables
Complete de-anonymisation of South Korean Resident Registration Numbers [240]	2015	23,163	Encrypted RRN	Yes

Table 2.1: Re-identification attack on real datasets [67].

[a] https://medium.com/@vijayp/of-taxis-and-rainbows-f6bc289679a1

2.3.1 Attributes

Assume a tabular dataset T, consist of records t, where each record corresponds to a different entity. The classification of the attributes of each record fall into four main categories:

- **Explicit Identifiers** are attributes that can identify uniquely a person. For example the Social Security Number.
- **Quasi-Identifiers** are publicly known characteristics of individuals. A Quasi-Identifier $\mathcal{QI}$ cannot be used to uniquely identify a person by itself but when combined with other quasi-identifiers it can lead to re-identification of a person. Either by narrowing down the possible identities to a small subset and this improves the confidence of an adversary regarding the possible real identity behind an anonymised record or in the worst case to pinpoint to a specific record. Examples of $\mathcal{QI}$ attributes are Gender, Zip Code, and Date of Birth.
- **Sensitive Attributes** - SA are the fields that an adversary wishes to know. Examples of SA can be the Disease or the Salary of a person in a medical or financial dataset respectively. Usually, there is only one SA in the table, but that is not always the case [141, 286, 276, 144, 154].
- **Non-Sensitive Attributes** Any attributes that does not fall in any the previous mentioned categories

An example of a table with Explicit Identifiers, Quasi-Identifiers and Sensitive Attributes is illustrated in Table 3.1.

EI		QI		SA
SSN	Gender	Age	Zip Code	Disease
987-65-4329	M	30	13000	Flu
987-65-4320	M	34	13500	HIV
987-65-4326	F	36	13500	Cancer
987-65-4322	F	38	16400	HIV

Table 2.2: Attribute Classification - Example.

2.3.2 Actors

A typical anonymisation scenario involves the following actors, as depicted in Figure 2.2 :

- **The Data Holder/Publisher:** The person or the organisation that wish to release a dataset with strong privacy guarantees. Usually, the data holder is also and the data publisher. In the case when the data holder cannot perform the anonymisation process because of lack of knowledge and/or limited resources then the role of data holder and data publisher are distinct.
- **The Record Owners:** Every entity which participates in the released dataset with one or more records
- **The Data Recipient:** Anyone with access to the published anonymous dataset.
- **The Adversary:** A malicious attacker or a nosy data recipient that wishes to gain additional knowledge about the SA of Record Owners.

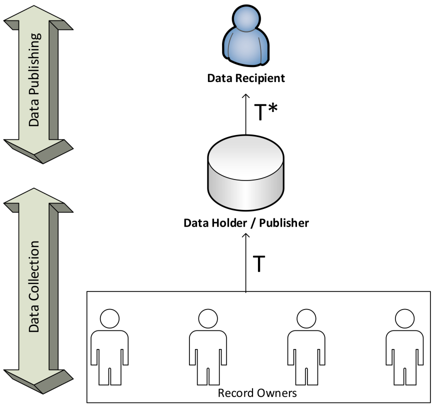


Figure 2.2: Roles in Data Collection/Publishing.

2.3.3 Single and Multiple-release publishing

The different data recipients may have different requirements on the published data. There are three main publishing scenarios and categorised as follows:

2.3.3.1 Publishing a Single release

The first and most common scenario assumes that the Data Publisher based on the privacy guarantees he desires, release the anonymous dataset only once. The original table T, or any subset of it, has never published previously and no further publication of T,or any subset of it, is going to be published after this T^* anonymised dataset.

2.3.3.2 Publishing releases in parallel

In the *Parallel releases* [285, 16, 124] scenario the original dataset T is released in a number of different T_i^* anonymised datasets. Depending on the different requirements of the data recipients, each T_i^* could consist of a subset of the original attributes. The goal of parallel releases is to reduce the information loss of the anonymous dataset that caused by the curse of dimensionality. Data Recipients may not have any use for specific QIs while requires other QIs to have the least possible changes in the anonymisation process. The Data Publisher should take into consideration that is possible the Data Recipients collude and try to combine the all the T_i^* available to them to gain more information of the Record Holders.

2.3.3.3 Publishing releases in sequence

Sequential publishing [263, 279, 30, 194, 33, 32, 18, 244, 260, 219, 218] refers to the incremental publication of anonymised dataset. Assume that a company periodically release anonymised data about its customers. The data on T are expected to change over time, usually with the addition of new records and alteration or deletion of existing records. The Data Publisher should not ignore the already published datasets, as the anonymity of a record owner is at risk when an adversary is cross-examining multiple releases.

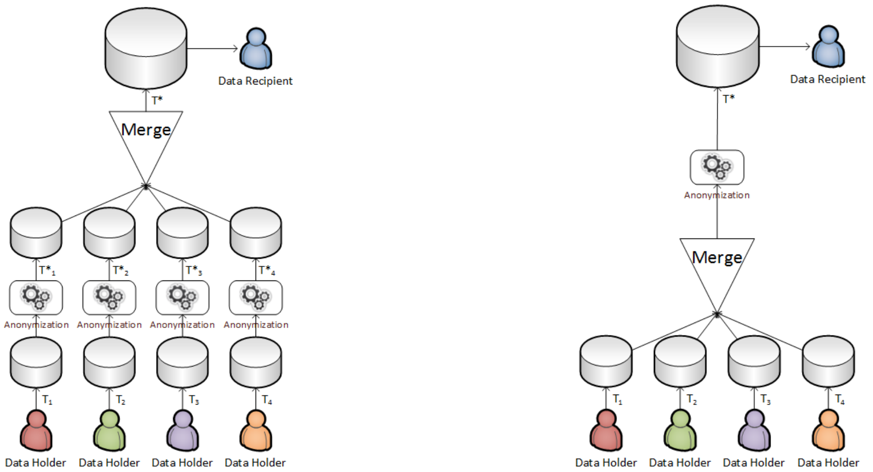
2.3.4 Centralised vs decentralised data publishing

The main focus in literature is on the centralised data publishing where a data publisher who possess the entire dataset from one or multiple data holders. For example, all of the countryś hospitals can send their datasets to the countryś ministry of health. A hospital does not have to anonymise their dataset and trust the ministry of health performs the anonymisation process in the entire collection of datasets. The case of decentralised publishing is when the data are distributed among multiple data holders who do not trust each other, but the data recipients wish to analyse the union of their datasets. However, due to legal or commercial constraints, the data holders do not wish to give access to raw data to anyone else[264, 96, 163, 223]. In this context, data might be partitioned among various parties in different ways [197, 105, 119, 283, 284]:

- Vertical partitioning (VP): data holders have disjoint sets of attributes but with the same individuals.
- Horizontal partitioning (HP): data holders have disjoint sets of individuals but with the same attributes.
- Arbitrary partitioning (AP): it is an intermix between VP and HP. This is the most realistic scheme in which datasets may contain an undetermined number of coincident attributes and individuals.

In the case of multiple data holders who want to publish their data in a common anonymised table T^* there is two different approaches, as illustrated in Figure 2.3, and they are the following:

- **Anonymise-and-Aggregate:** Each data holder anonymise the data independently and then aggregate all their anonymous tables to one and release it to data recipients [222]. While in this approach the cost is relatively manageable and the solution is very fast, it has the disadvantage to introduce unnecessary degradation of the data utility.



(a) Anonymise and Aggregate.

(b) Aggregate and Anonymise.

Figure 2.3: Collaborative data publishing methods.

- **Aggregate-and-Anonymise:** The second approach is more appropriate as it succeeds less data distortion for the same privacy guarantees as the first one. The process starts with the data holders aggregating all their original data and afterwards applying the anonymisation process. Since legal issues may prohibit the sharing of plain data with others more complex solutions as the introduction of a semi-trusted third party or Secure Multi-party Computation [126, 122] protocols can be used. This does not come without cost since a significant computational overhead is added for ensuring the encrypted communication without leakages.

2.4 Data Transformation Techniques

The procedure of transforming the original dataset T to T^*, see Figure 2.2, with respect to the privacy requirements of the data holder is called anonymisation. The original dataset T undergo various data transformation to achieve the desired properties in the anonymous dataset T^*. In what follows, the basic data transformations techniques that are employed in anonymisation methods are introduced.

2.4.1 Generalisation

The transformation of generalisation is the replacement of a value of a $\mathcal{QI}$ with a different one, which is an abstraction of the original. This can be achieved through the *Domain Generalisation Hierarchy* (DGH), which is a lattice or a graph that represent the solution space for the anonymisation problem. Every node in the lattice is a different combination of generalisation levels of the QIs. Nevertheless, in quantitative data is possible the DGH to be dynamically defined by the anonymisation algorithm as shown in [94]. The values of the QIs are usually replaced based on predefined generalisation hierarchies, which dictates how each value of the original domain is replaced by a more generic one. An example of a hierarchy is depicted in Figure 2.4.

Example 2.4.1. In Figure 2.4 the attribute Age when it is not generalised, at level 0, has six distinct values. Moving up to level 1, the first two values 12 and 24 are mapped to (<25), the values 34 and 48 falls within the range (25<...<60) and finally the values 65 and 88 mapped to (>60). At level 2 the three different values of level 1 is now mapped to a more general value which now includes all the domain of attribute Age and represented by *. The attribute Gender, since only have two distinct values, at level 1 the values Male and Female should be generalised to a more general value e.g. human, or *. The DGH is now constructed with the three levels of generalisation of attribute Age and the two of the attribute Gender as illustrated in Figure 4.1.

The generalisation transformations can be distinguished into two major categories:

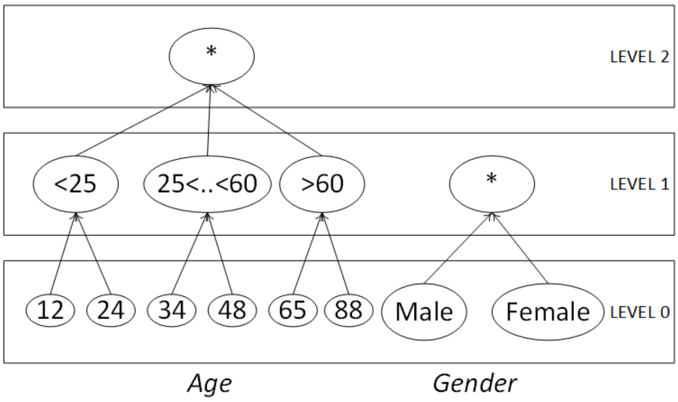


Figure 2.4: Taxonomy Tree.

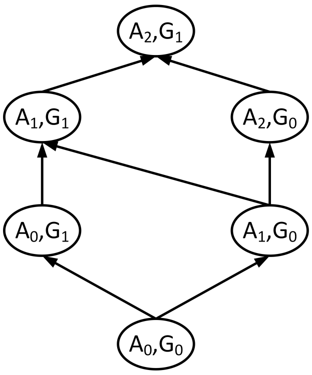
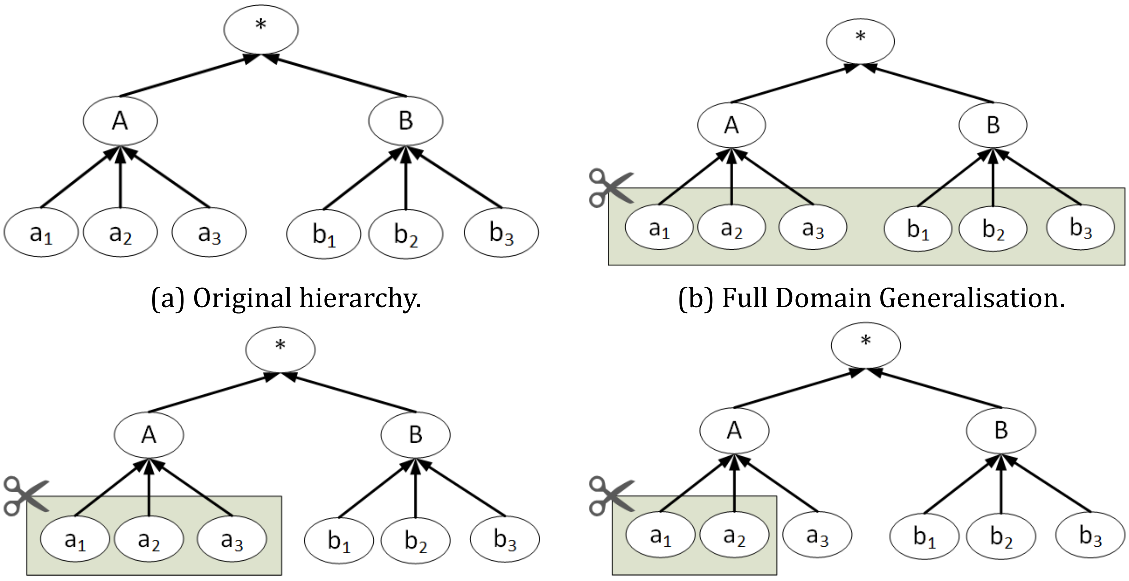


Figure 2.5: Domain Generalisation Hierarchy of attributes Age and Gender.

1. **Global recoding** refers to the global substitution of an attribute value with a more generic one. For example, if a value a_1 is replaced by value A, then all the instances of a_1 in T, will be replaced by A. There are three approaches to performing global recoding.

   - **Full domain generalisation:** All attribute values are generalised to the same generalisation hierarchy level [132, 211, 237, 238, 235, 125, 66]. This method has the advantage to present the anonymous dataset T^* at the same granularity level for all its data, making it easier for the human eye to group, read and understand.On the other hand, original values might be unnecessarily generalised.
   - **Subtree generalisation:** In this case, all sibling nodes are required to be generalised to the same value. The nodes of the other subtrees are independently generalised if that it is necessary [19, 112, 85, 86, 37].

- **Sibling generalisation:** In this approach, some of the sibling nodes are generalised, while others remain intact. This provides greater flexibility to anonymisation algorithm to reduce the information loss with fewer generalisations. However, the computational cost is increased [132] since the solution's search space is expanded.



(a) Original hierarchy.

(b) Full Domain Generalisation.

(c) Subtree Generalisation.

(d) Sibling Generalisation.

Figure 2.6: Global Recoding Subcategories.

Example 2.4.2. In full domain generalisation,as seen in Figure 2.6b, if a_1,a_2 and a_3 are generalised to A, then equivalently, b_1,b_2 and b_3 are also generalised to B. In the case of subtree generalisation, Figure 2.6c, when a_1 generalised to A, then a_2 and a_3 also have to generalised to A, but b_1, b_2 and b_3 can remain unchanged in T^*. As depicted in Figure 2.6d, the sibling generalisation, the values a_1 and a_2 can be generalised to A without requiring a_3 to be similarly generalised.

2. **Local recoding**: Local recoding allows only some appearances of a value to be generalised. This means $\mathcal{QI}s$ values can be generalised to different levels across the anonymised dataset.

   - **Cell generalisation:** The Cell generalisation allows the generalisation of an instance of the value while the other instances are not affected by this operation. For example, the value 13500 of the attribute Zip Code can be generalised to 135** while another instance of the same value to generalised to 13*** in order to provide the privacy guarantees.

- **Multidimensional generalisation:** A multidimensional generalisation can be obtained by applying a single function to a relation, which has several QIs and their taxonomy trees, by generalising $QI = (v_1, ..., v_n)$ to $QI^* = (u_1, ..., u_n)$ in a way that $v_i = u_i$ or v_i to be a descendant node of u_i in the taxonomy of attribute i [133, 89]. It is the same operator as the cell generalisation with the difference that it takes into account multiple dimensions of the records.

Example 2.4.3. From the Table 2.3a and for the QIs Age and Zipcode a single dimensional cut occurs at the Zipcode dimension at 10711. As is depicted in Figure 2.7(b), after this cut there are no more possible cuts from side to side since any cut on the Age axis would create areas with only one member which means this member's privacy is breached. Multidimensional partitioning, Figure 2.7(c), would allow one more cut inside the left region at the value 36 on Age axis. All regions now have more than one member.

In the case of Single dimensional partitioning two groups are formed

(a) $[35 - 38], [10710 - 10711]$

(b) $[35 - 38], [10712]$

In Multidimensional partitioning, there are three groups formed. The data utility is increased when smaller and more precise groups are formed. These groups are:

(a) $[35 - 37], [10712]$

(b) $[35 - 36], [10710 - 10711]$

(c) $[37 - 38], [10710 - 10711]$

When the Data Publisher have to make a decision on the generalisation operator there are two factors he must have in his mind. The first is the quality of the generalisation; the more flexible an operator is, the smaller the information loss will be. The second factor is the computational cost of the algorithm. The trade-off here is that more flexible operators offer a greater solution space which makes the computational cost to rise.

Age	Gender	Zipcode	Disease
35	Male	10711	Flu
35	Female	10712	Viral Infection
36	Male	10711	Heart Disease
37	Male	10710	HIV
36	Female	10712	Mastitis
38	Male	10711	Prostate Cancer

(a) Original Table.

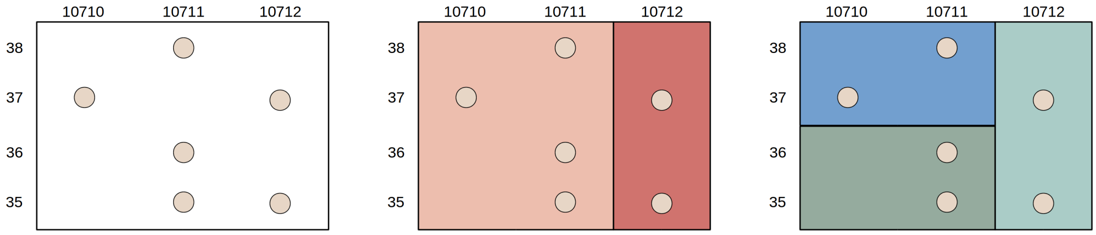
Age	Gender	Zipcode	Disease
35-38	Male	10710 - 10711	Flu
35-38	Female	10712	Viral Infection
35-38	Male	10710 - 10711	Heart Disease
35-38	Male	10710 - 10711	HIV
35-38	Female	10712	Mastitis
35-38	Male	10710 - 10711	Prostate Cancer

(b) Single-dimensional anonymisation.

Age	Gender	Zipcode	Disease
35-36	Male	10711	Flu
35-37	Female	10712	Viral Infection
35-36	Male	10711	Heart Disease
37-38	Male	10710 - 10711	HIV
35-37	Female	10712	Mastitis
37-38	Male	10710 - 10711	Prostate Cancer

(c) Multidimensional anonymisation.

Table 2.3: Single vs Multidimensional anonymisation.



(a) Patients table.

(b) Single-dimensional partitioning.

(c) Strict multidimensional partitioning.

Figure 2.7: Single-dimensional vs Multidimensional - Spatial representation for QIs Zipcode and Age.

2.4.2 Suppression

The removal of certain records or values from the original dataset in order to anonymise it is called Suppression. The are three categories of suppression. The *Tuple or Record Suppression* [237, 19], where the entire record is suppressed. The *Value Suppression* [266]with the suppression of a given value throughout the entire table. Finally, the *Cell suppression* [160] suppresses only some instances of a value in the table.

Example 2.4.4. In this example, a slightly modified version of the Table 2.3a is used to show how each suppression method works, resulting tables with no distinct records.

- **Record Suppression** On the Table 2.4a the third record has been deleted and now there is no need for further suppression or generalisation since the remaining records form two groups with the same $\mathcal{QI}s$.

- **Value Suppression** Both values M and F of the $\mathcal{QI}$ Gender has been suppressed from the Table 2.4b

- **Cell Suppression** In the Table 2.4c the value of a single cell has been suppressed. The suppressed value works like a wildcard that in this example could either "hide" value 10710 to match the first record, or value 10711, to match the fifth record.

2.4.3 Bucketisation

When data recipients demand the original values of QIs then the data publisher may use the transformation of *bucketisation*, often referred to as Anatomisation [277], to break the association between $\mathcal{QI}s$ and the SA. This can be done simply by publishing

Age	**Gender**	**Zipcode**	**Disease**
35	Male	10711	Flu
36	Female	10712	Viral Infection
~~58~~	~~Male~~	~~15711~~	~~Heart Disease~~
36	Female	10712	Mastitis
35	Male	10711	Prostate Cancer

(a) Record Suppression.

Age	**Gender**	**Zipcode**	**Disease**
35	*	10711	Flu
36	*	10712	Viral Infection
35	*	10712	Heart Disease
36	*	10712	Mastitis
35	*	10711	Prostate Cancer

(b) Value Suppression.

Age	**Gender**	**Zipcode**	**Disease**
35	Male	10710	Flu
36	Female	10712	Viral Infection
35	Male	*	Heart Disease
36	Female	10712	Mastitis
35	Male	10711	Prostate Cancer

(c) Cell Suppression.

Table 2.4: Suppression example.

them in separate tables. The published tables share a common attribute, the $group - id$. Now from the SA table, any SA value with $group - id = i$ can be linked to any individual with $group{-}id = i$ at the $\mathcal{QI}s$ table. By publishing these tables individually, the connection between $\mathcal{QI}s$ and the SA is broken and without the need for generalise or suppress any QIs value. Compared to the generalisation approach, this approach introduces a different type of information loss and gives more accurate answers to aggregation queries that involve $\mathcal{QI}$s since the original values remain intact.

Example 2.4.5. By splitting the Table 2.5a in two fragments with the bucketisation method, the result is the $\mathcal{QI}$ Table 2.5b and the SA Table 3.8. An adversary, with the original $\mathcal{QI}$s values on his hands may be able to find his target on the Table 2.5b but he cannot link that individual to a certain disease. The non-malicious data recipient with the original $\mathcal{QI}$s values have the advantage to be able to analyse the data in more detail.

There is a drawback to the bucketisation method. The adversary could infer the participation or not of his target in a released dataset. Sometimes only the participation in a table could be considered sensitive. For example, if an adversary could infer the participation of his target in a released table with Sexually Transmitted Diseases, even without knowing the exact disease this is a serious privacy breach.

2.4.3.1 Slicing

The basic idea behind Slicing [140] is the breaking of associations across columns while preserving associations within each column. Columns can be formed with one or more $\mathcal{QI}$s, SA or both. The utility of the dataset is preserved by forming groups with highly correlated attributes, while the breaking of the associations between uncorrelated attributes enhances the privacy protection. Note that the Bucketisation/Anatomisation transformation can be considered as a special case of Slicing, with only two columns, the first containing all the $\mathcal{QI}$s and the other only the SA.

2.4.3.2 Disassociation

Disassociation is an anonymisation transformation proposed by Terrovitis et al. [247]. This technique provides protection against identity disclosure on sparse multidimensional data without suppressing or generalising the original terms. Disassociation partitions the original records into smaller and disassociated subrecords. The goal is to hide infrequent term combinations in the original records by scattering terms in disassociated sub-records, as shown in Figure 2.8.

Age	Gender	ZipCode	Disease (sensitive)
40	Male	50100	Hepatitis
40	Male	50100	Hepatitis
40	Male	50200	HIV
42	Male	50200	Hepatitis
42	Male	50200	HIV
44	Male	50300	HIV
46	Female	50400	Flu
48	Female	50300	Flu
48	Female	50300	Cancer
48	Female	50400	Cancer

(a) Original Data.

Age	Gender	ZipCode	GroupID
40	Male	50100	1
40	Male	50100	1
40	Male	50200	1
42	Male	50200	1
42	Male	50200	1
44	Male	50300	1
46	Female	50400	2
48	Female	50300	2
48	Female	50300	2
48	Female	50400	2

(b) QI Table.

GroupID	Disease (sensitive)	Count
1	Hepatitis	3
1	HIV	3
2	Flu	2
2	Cancer	2

(c) Sensitive data Table.

Table 2.5: Bucketisation Example.

Age , Gender	ZipCode, Disease
(40 , Male)	(50100 , Hepatitis)
(40 , Male)	(50100 , Hepatitis)
(40 , Male)	(50200 , HIV)
(42 , Male)	(50200 , Hepatitis)
(42 , Male)	(50200 , HIV)
(44 , Male)	(50300 , HIV)
(46 , Female)	(50400 , Flu)
(48 , Female)	(50300 , Flu)
(48 , Female)	(50300 , Cancer)
(48 , Female)	(50400 , Cancer)

Table 2.6: Slicing.

ID	Records
r_1	itunes, flu, madonna, ikea, ruby
r_2	madonna, flu, viagra, ruby, audi a4, sony tv
r_3	itunes, madonna, audi a4, ikea, sony tv
r_4	itunes, flu, viagra
r_5	itunes, flu, madonna, audi a4, sony tv
r_6	madonna, digital camera, panic disorder, playboy
r_7	iphone sdk, madonna, ikea, ruby
r_8	iphone sdk, digital camera, madonna, playboy
r_9	iphone sdk, digital camera, panic disorder
r_{10}	iphone sdk, digital camera, madonna, ikea, ruby

(a) Original Dataset.

	Record chunks		Term chunk
	C_1	C_2	C_r
r_1	itunes, flu, madonna		
r_2	flu, madonna	audi a4, sony tv	ikea,viagra
r_3	itunes, madonna	audi a4, sony tv	ruby
r_4	itunes, flu,		
r_5	itunes, flu, madonna	audi a4, sony tv	

(b) Cluster P_1, $|P_1| = 5$.

	Record chunk	Term chunk
	C_1	C_r
r_6	madonna, digital camera	
r_7	iphone sdk, madonna	panic disorder
r_8	iphone sdk, digital camera, madonna	playboy, ikea ruby
r_9	iphone sdk, digital camera	
r_{10}	iphone sdk, digital camera, madonna	

(c) Cluster P_2, $|P_2| = 5$.

Figure 2.8: Anonymisation by Disassociation.

2.4.4 Permutation

Based on the concept of anatomisation Zhang et al. [291] introduced the permutation method. This method aims to anonymise tables with numerical SA. The first step is to group the data records based on their $\mathcal{QI}$ and then in each group to shuffle the SA values. Aggregate queries can be answered more accurate with the permutation than suppression and/or generalisation-based techniques.

Example 2.4.6. A naive example of permutation is presented with the dataset of Table 2.7a and the permuted version of it, the Table 2.7b.

Assuming that the query "What is the average salary of people younger than 42 years old ?" needs to be answered, on the original Table 2.7a the answer is equal to

48250. On a generalisation-based approach a truthful hypothesis could be that the intervals of $\mathcal{QI}$ Age are [30-40] and [41-50].In this case, the result of the same query would include the average of all the records and would be equal to 53000. However, using permutation the result to the query in this instance would be 50750, which is more accurate since the average would be computed based only on the 4 records that match the query.

Name	Age	ZipCode	Gender	Salary
Achilleas	30	50110	M	45000
Bill	38	50120	M	46000
Carlos	40	50130	M	47000
Debby	41	50220	F	55000
Emily	43	50260	F	65000
Fred	50	50240	M	60000

(a) Original Table.

Group	Age	ZipCode	Gender	Salary
1	40	50130	M	45000
1	38	50120	M	46000
1	30	50110	M	47000
2	50	50240	M	55000
2	41	50220	F	65000
2	43	50260	F	60000

(b) Permuted Table.

Table 2.7: Permutation Example.

Permutation may seem like an efficient method but randomly permuting SA values may lead to privacy breaches because logical links exist between the different attributes.

Example 2.4.7. This weakness of permutation is illustrated in Table 2.8 where the random permutation of data is problematic. The adversary can infer that the CEO is more likely to have the highest income while the Unemployed the lowest one, therefore, exposing their SA values.

2.4.5 Perturbation

The original dataset is replaced by a synthetic one in order to keep the distortion of statistical information values to acceptable levels. Since the records do not correspond

Age	Gender	Job	Income (permuted)
44	M	Engineer	70000
44	M	CEO	5000
44	M	Unemployed	43000
35	M	Engineer	100000
35	M	Manager	45000

Table 2.8: A Problematic Permuted Table.

to their original values after the perturbation, the adversary cannot successfully link a record to an individual. Perturbation methods can be further categorised as follows:

- **Noise addition** On a table T and in a numerical SA with the value v_i, Noise addition T adds a random number r that follows a distribution. The anonymised value is the $v_i + r$. [26]

Age	Gender	Salary α	$r = [-5000, 5000]$	Salary $\alpha + r$
35	M	45000	4325	49325
37	M	50000	-2751	47249
25	F	38000	-4198	33802
33	M	40000	-3706	36294
40	F	36000	2136	38136
26	M	67000	2524	69524
60	F	78000	-4616	73384
45	F	80000	-1614	78386
	Sum	434000		426100
	Mean	54250		53262,5
	Standard deviation	18100,90763		17918,94265
	Variance	327642857,1		321088505,7

Table 2.9: Noise Addition

Example 2.4.8. This example illustrates how noise addition works on attribute Salary on Table 2.9. In the released table the attribute Salary will be replaced with the last column of Table 2.9. This column is the sum of the original yearly salary with a random value $v \in [-5000, 5000]$. At the bottom of the table, specific aggregate functions are shown, such as the sum and standard deviation, which can efficiently be computed with inconsiderable errors in the reported values.

Age	Gender	Zipcode	Disease
35	Male	10711	Flu
36	Female	10712	Viral Infection
58	Male	15711	Heart Disease
46	Female	10712	Mastitis
25	Male	10711	Prostate Cancer

(a) Original Data.

Age	Gender	Zipcode	Disease
35	Male	10711	Viral Infection
36	Female	10712	Mastitis
58	Male	15711	Prostate Cancer
46	Female	10712	Heart Disease
25	Male	10711	Flu

(b) Data Swapping .

Table 2.10: Data Swapping Example

- **Data swapping** Data swapping is used for both numerical and categorical SA. The exchange of SA values among the records anonymises the table [73, 255, 174].

  Example 2.4.9. The values of the SA in Table 2.10a are swapped randomly to Table 2.10b while the $\mathcal{QI}s$ remains intact.

  This result can be considered satisfactory but that is not always the case. The SA values may have interdependencies which can lead to inconsistencies on the released table as in the case of gender-specific diseases. For instance in a medical dataset, assigning $Prostate\ cancer$ to a Female or $Mastitis$ to a Male are impossible combinations, nonetheless, this can happen with random swaps.

- **Synthetic data generation** Based on the original data, Synthetic data generation builds a mathematical model and uses it to generate the anonymised table with synthetic records in such manner that basic statistical measures or relationships are preserved [208, 157, 5, 12]. The major drawback in this approach is that these data are no longer useful for analysis on random subdomains. To reduce the impact of this problem two approaches emerged, the Partially synthetic [204] and the Hybrid data [57, 165, 178]

Data Transformation	Data Privacy	Utility
Bucketisation	Low	High
Generalisation	Average	Average
Permutation	Average	Average
Perturbation	High	Low
Suppression	High	Low

Table 2.11: Data Transformation: Data Privacy vs Utility.

- **Microaggregation** is a perturbation method of aggregating values of attributes to reduce re-identification risk. This method this method has two steps, the first is the *data partitioning* and the second is the *partition aggregation* [58]. The data partitioning breaks the dataset T in subsets $T_{s_1}, T_{s_2}, ...T_{s_n}$ in such way that for $i \neq j$, $T_{s_i} \cap T_{s_j} = \emptyset$ and $T_{s_1} \cup T_{s_2} \cup .... \cup T_{s_n} = T$. Then on the partition aggregation step, a representative value for each cluster is selected, which is usually the median or the mean value. The original values on the clusters are replaced with the corresponding representative value. A parameter k is defined to control the minimum size of each cluster.

  The data publisher should be aware for dependencies among the attributes. For example in a hypothetical table that contains the attributes "Hours paid for" , "Wage Rate" and "Wage Sum". The multiplication of the first two should result the "Wage Sum". While the mathematics is valid in the original data, in the microaggregated version the function is no longer valid. To avoid this problems, some solutions have been proposed [251, 34] that inserts constraints which microaggregation algorithms should not violate. Originally, microaggregation was focused on numerical attributes, it was later extended to cover categorical attributes as well [250].

The trade-off between privacy and utility is crucial [27, 139, 213] to data anonymisation. Very strong privacy guarantees usually lead to an anonymised table with no practical use while minimizing the information loss make the adversary more threatening. The Table 2.11 it is an overview of the impact of each data transformation on privacy and utility.

2.5 Information Metrics

As previously mentioned the information loss plays a significant role in the anonymisation process. The major concern of data publisher and data recipient is to find the best possible balance between privacy and data utility. To measure the data utility of an anonymised table T^* various information metrics have been proposed and can be categorised according to their purpose into three major categories, namely: *general, special* and *trade-off purpose*. More precisely:

- **General Purpose:** Since the data publisher does not know the analysis that the data will undergo by data recipient, he wants the released dataset T^* to be as similar as possible to the original table T. To this end, the *principle of minimal distortion* (MD) was introduced in [211, 235, 238] which simply issues a penalty for every QI value that is generalised. This metric does not take into account the importance of each QI attribute neither the size of the generalisation hierarchy. Xiao and Tao [278] proposed a slightly more complicated metric called *ILoss*. *ILoss* assigns to each cell on the table a number between 0 and 1, which is proportional to the size of the hierarchy and the position of that value in the hierarchy. Zero is attributed to cells that with no generalisation, and one when there is total generalisation or suppression. Moreover, the data publisher could set weights to the attributes to reflect their importance.

  MD and $ILoss$ charge a penalty no matter which record is generalised and does not take into consideration the other values in the table T^*. In some cases that could be problematic as shown in the next example.

  Example 2.5.1. In Figure 2.6 if there is 9 values of a_1 and only 1 value of a_2, by generalising both a_1 and a_2 values to A this would charge a penalty equal to 10 values for being generalised. The same happens if the a_1 has 5 values and the a_2 also 5 values. Both cases have the same penalty but there is a difference. Obviously, in the first case the 9 records were already indistinguishable, while in the second case only 5 records were indistinguishable. This means that in the second case more originally distinguishable records became indistinguishable.

  Skowron and Rauszer [221], introduce the *discernibility metric (DM)* in which the charged penalty for each value depends on other values in the release.When a record belongs to a group of size s, the penalty for being indistinguishable from other records will be equal to s if it is generalised.

The *Ambiguity Metric (AM)* [170] was designed for k-anonymity framework, which will be discussed later on. For every record r in the anonymised table T^*, AM calculates the number of records in T that could be have been generalised as r. This number is the ambiguity of r. The AM for the T^* is defined as the average ambiguity of all records in T^*. A drawback of the AM metric is that it counts also combinations of attribute values that do not appear in the original table.

- **Specific Purpose:**

  In the rare occasion when it is known in advance the analysis that the data recipients wish to perform on the anonymised dataset, then this information can be taken into consideration during the *anonymisation process* by the data publisher. Some may argue that if the analysis is known then the data publisher could perform the data mining process and provide only the results without releasing an anonymised table. This is impractical for a non-expert data publisher and undesirable for the data recipient.

  Generalisation and suppression can influence the process of data mining either in a positive or a negative way. For example, generalisation in some cases could help the data mining algorithm to perform better with the generalisation of over-specialised attributes. The *classification metric (CM)* introduced by Iyengar [112] to measure the classification error on the training data. When a record is suppressed or generalised to a group in which the record's class is not the majority class then the CM metric charges a penalty.

  The intuition is that a record having a non-majority class in a group will be classified as the majority class, which is an error because it disagrees with the recordś original class.

  As a data metric though, CM does not address the problem of over-specialisation of values efficiently.

- **Trade-off Purpose:** As the name indicates the trade-off metrics try to balance the privacy and the information requirements, at every anonymisation operation of the anonymisation algorithm.

  For example, an anonymisation process that iteratively specialises a general value into child values, in each specialisation s will gain in terms of information, denoted as $IG(s)$ and lose in terms of privacy, $PL(s)$. The metric proposed by Fung et al. [85] prefers the specialisation s that maximises the information gain for each loss of privacy:

$$IGPL(s) = \frac{IG(s)}{PL(s) + 1}$$

This is a generic model and the choice of $IG(s)$ and $PL(s)$ depends on the information metric and the privacy model.

- **KL-Divergence:** All the aforementioned metrics fails to take into consideration the distribution of the attribute values. For example, if the values of QI Age is uniformly distributed and independent from other QI then replacing it with a range of values would have a little effect since the data analyst, based on the principle of maximum entropy can assume a uniform distribution within the range. However, when the Age distribution is skewed the uniformity assumption could lead the analyst to false results. The *Kullback-Leibler divergenhe ce* or commonly known as *KL-divergence* can be used to overcome this problem. In the anonymasation scenario, the original table T and the anonymous T^* are treated as probabilities of distribution p_1 and p_2. The *KL-divergence* between these is defined as:

$$KL = \sum_t p_1(t) \log \frac{p_1(t)}{p_2(t)}$$

Similar to *KL-divergence* that measure the distance between original and the probability distribution reconstructed from the anonymous data are the L_p*-Norm* [9], *Hellinger Distance*. For more details, see [40].

2.6 Attacks and countermeasures

2.6.1 Attacks On Anonymised Publications

On the anonymisation domain, there are various attack scenarios which include adversaries with different strengths and goals. The goals of an adversary vary from the complete identification of the record holders to simply learn if the record of his target is included or not to a published dataset. The different scenarios can be classified into the following categories:

2.6.1.1 Record Linkage:

The goal of *record linkage* is to link successfully a record of the anonymised database to an individual. Using his background knowledge of the QIs an adversary could link a small group of records or a unique record to an individual. In addition to the Sweeny's [238] attack that mentioned earlier, the example that follows will illustrate that a re-identification technique can be as simple as a database "JOIN".

Example 2.6.1. In this example it is assumed that there are two datasets, the first one, denoted as $Dataset_1$, does not have any SA attribute in it, therefore and anyone can access it and there is no need to anonymise it. Such dataset is the voters list where the explicit identifiers and some $\mathcal{QI}s$ are made available to the adversary. The second is an anonymised dataset, denoted as $Dataset_2$, e.g. medical records without explicit identifiers, but with the SA values and $\mathcal{QI}s$. The adversary can perform a matching on $\mathcal{QI}s$ from both datasets and he expects to get as result a unique re-identification or a small group of possible matches. When an adversary targets one or some specific individuals and a dataset like $Dataset_1$ is not available to him, then he can use his background knowledge about the $\mathcal{QI}s$ of his victims to perform the attack. In Table 2.12, if the adversary knows the $\mathcal{QI}s$ values of Cher, he can easily infer that she has HIV since is the only one with that $\mathcal{QI}s$ values.

Name	Age	Gender	Zipcode
Achilleas	23	Male	11527
Brooke	44	Female	11045
Cher	22	Female	15345
David	33	Male	50100
Eve	54	Female	50102

(a) Dataset_1.

Age	Gender	Zipcode	Disease
54	Female	50102	Mastitis
44	Female	11045	Uterine Cancer
33	Male	50100	Prostate Cancer
22	Female	15345	HIV
81	Female	11530	Alzheimer
23	Male	11527	Flu
29	Male	50100	HIV
40	Female	11000	Uterine Cancer

(b) Dataset_2.

Table 2.12: Record linkage.

2.6.1.2 Attribute Linkage

The *attribute linkage attack* refers to the scenario when it is not possible for an adversary to link a specific record to his target but he is able to link a specific SA value. For example, in Table 2.12b the adversary may know that his victim Brooke, a female around 40 to 45 years old which lives in an area where the Zip Codes starts with 110. While he cannot conclude which record is linked to his victim, nevertheless he can infer which SA value she has, in our example "Uterine Cancer". It becomes apparent that the lack of enough diversity on the sensitive values of each group makes this attack successful.

When groups are formatted based on $\mathcal{QI}$s, the adversary could infer the sensitive value of a person even if he cannot point which record belongs to that person.

2.6.1.3 Table Linkage Attack:

On the two aforementioned attacks, Record and Attribute linkage, it is assumed that the adversary already knows that his target record is on the released anonymous dataset table. However, that is not always the case. Sometimes the presence or the absence of an individual from an anonymous dataset reveals sensitive information about him. If an adversary can confidently infer the presence or absence of a victim in that dataset, then he has successfully carried a *table linkage* attack [171, 172].

2.6.2 Countermeasures to Record Linkage

In what follows the most well-known countermeasures against record linkage are presented. Of specific interest is k-anonymity, one of the first and most used methods in the field and its variations.

2.6.2.1 k-anonymity

Samarati and Sweeney [212, 238] introduced the notion of k-anonymity as a countermeasure to record linkage. A dataset is called k-anonymous when it returns at least k records for any set of $\mathcal{QI}s$ values. The set of records with the same $\mathcal{QI}s$ is called an equivalence class (EC). From the adversary's perspective, when he knows the $\mathcal{QI}$ of a target individual, the probability to successfully link his target record is never greater than $\frac{1}{k}$. This probability takes into consideration that the adversary knows that his target is in the dataset and also knows all the possible values of $\mathcal{QI}s$ of the target.

A formal definition of k-anonymity is the following:

Definition 1 (k-anonymity)**.** A table T is k-anonymous if for every record $t \in T$ there exist $k-1$ other records $t_{i_1}, t_{i_2}, ..., t_{i_{k-1}} \in T$ such that $t[\mathcal{C}] = t_{i_1}[\mathcal{C}] = t_{i_2}[\mathcal{C}] = ... = t_{i_{k-1}}[\mathcal{C}]$, $\forall \mathcal{C} \in \mathcal{QI}$

Example 2.6.2. The original table T_1 in Table 2.13a, is transformed into T_1^* , see Table 2.13b, with the generalisation of the $\mathcal{QI}$ Age and Zip Code, and with generalisation to the maximum level of Nationality, which is basically the suppression of this $\mathcal{QI}$ attribute. Clearly, the anonymised table is 4-anonymous, since for every record there are at least three others with the same $\mathcal{QI}$ values.

Non-Sensitive			Sensitive
Zip Code	Age	Nationality	Condition
13053	28	Russian	Heart Disease
13068	29	American	Heart Disease
13068	21	Japanese	Viral Infection
13053	23	American	Viral Infection
14853	50	Indian	Cancer
14853	55	Russian	Heart Disease
14850	47	American	Viral Infection
14840	49	American	Viral Infection
13053	31	American	Cancer
13053	37	Indian	Cancer
13068	36	Japanese	Cancer
13068	35	American	Cancer

(a) T_1

Non-Sensitive			Sensitive
Zip Code	Age	Nationality	Condition
130**	< 30	*	Heart Disease
130**	< 30	*	Heart Disease
130**	< 30	*	Viral Infection
130**	< 30	*	Viral Infection
148**	> 40	*	Cancer
148**	> 40	*	Heart Disease
148**	> 40	*	Viral Infection
148**	> 40	*	Viral Infection
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer
130**	30-40	*	Cancer

(b) T_1^* 4-anonymous

Table 2.13: 4-anonymous

As Aggarwal [7] shown the *Curse of dimensionality* plays a crucial role in the anonymisation procedure. When the number of $\mathcal{QI}s$ becomes large it becomes more difficult to anonymise the dataset without an unacceptably high amount of information loss. Therefore, high-dimensional datasets suffer the most from the curse of dimensionality. To decrease the information loss, practitioners often anonymise high-dimensional datasets by using only a subset of the $\mathcal{QI}s$ depending on the data sharing purpose, and also have multiple parallel releases of the table with different subsets of $\mathcal{QI}$s.

Note that the data publisher may have different preference on the $\mathcal{QI}s$ [52] that are going to be generalised or suppressed, for example, in a medical dataset $\mathcal{QI}s$ such as age and profession may be more significant than the Zip code of the underlying patient.

k-anonymity though microaggregation Microaggregation to satisfy the k-anonymity property was first studied in [60, 59, 53]. The classic approach of microaggregation is the *univariate microaggregation*, which processes independently each numeric attribute and does not guarantee the k-anonymity property. The *multivariate microaggregation* refers in the case when the clustering process takes into account all the $\mathcal{QI}s$

and microaggregates these attributes together so the k-anonymity property is satisfied. Nonetheless, this approach suffers larger information loss when compared to *univariate microaggregation*.

k**-map** Sweeney [238] proposed the k-map property. Assume that T_{id} is an identification database which is k-anonymised to produce T^*_{id}. The data publisher wish to release a dataset T^* from the original T. To comply with k-map property each record in the disclosed k-map dataset T^* must be related to at least k records in T^*_{id}. The difference with k-anonymity is that k-anonymity takes into consideration the original dataset T when the algorithm forms the equivalence classes while the k-map the k-anonymous identification database T^*_{id}. This method can offer the same guarantees as k-anonymity and simultaneously reduces the information loss. However, a major drawback of k-map is that combining all the available external data with the data which the data holder wishes to release is a difficult task and not always feasible, rendering k-map unusable in real life scenarios.

(X,Y)-Anonymity One of the assumptions of k-anonymity is that each record holder has only one record in the dataset. However there are cases where this assumption does not hold. For example datasets of medical records can have more than one record per individual. Assume that there is a dataset with a set of $\mathcal{QI}$s Age, Gender and Zip Code, with SA Disease and explicit identifier the Social Security Number (SSN). An individual can have more that one disease, therefore, more than one records. The SSN as an explicit identifier is removed and the k-anonymous released dataset may have at least k records per each EC but there is no guarantee that contains at least k distinct individuals. In the extreme case that a record holder has k records, then an EC could contain only records from one individual. To tackle this problem Wang and Fung [263] introduced the notion of *(X,Y)-Anonymity*. To satisfy *(X,Y)-Anonymity* each value on X must be linked to at least k distinct values on Y. In this example Y={SSN} and X={Age, Gender, Zip Code}. Note that Y could also be the SA disease so each group is associated with a diverse set of SA values, making it even more difficult to infer a SA value.

2.6.2.2 (1,k)-anonymisation , (k,1)-anonymisation, (k,k)-anonymisation

A relaxation of k-anonymity proposed by Gionis et al [92]. They introduced the notions of (1,k)-anonymity and (k,1)-anonymity.

(1,k)-anonymity When the adversaries are aware only of the public available information about their targets, gathered in the table T_{pub} then the data publisher instead of performing k-anonymisation, he could generalise the table entries in such way that the public data T_{pub} of every individual are consistent with at least k records of the released table T^*. (1,k)-anonymity is similar to k-map. Every k-anonymous table is also a (1,k)-anonymised table but the converse is not necessarily true.

(k,1)-anonymity A table T^* is (k,1)-anonymous when any record in the table T_{pub} is consistent with at least k records on the original table T. As before a k-anonymous table is also (k,1)-anonymous.

Both of them offer a weaker protection of privacy when are compared to k-anonymity but its meant to used combined.

(k,k)-anonymity When an anonymous table satisfies both (k,1)-anonymity and (1,k)-anonymity then is a (k,k)-anonymous table. (k,k)-anonymous tables offer similar protection to k-anonymous tables when the attack scenario is an adversary who has only full knowledge on some of the individuals in the table. Using (k,k)-anonymity the data publisher may see higher utility compared to k-anonymity.

2.6.2.3 Non-homogeneous Generalisation

Researchers proposed [275, 245, 230] a more complex approach to reduce further more the information loss by not having the same generalised values of $\mathcal{QI}s$ within an EC with more than k members.

Example 2.6.3. From the original dataset in Table 2.14a the 2-anonymous dataset in Table 2.14b with homogeneous generalisation is produced. In Table 2.14c is a 2-anonymous non-homogeneous version of Table 2.14a which have different intervals, different level of generalisation and suppressed values of in the $\mathcal{QI}s$ for the first 3 records. Assuming that the adversary knows all the $\mathcal{QI}s$ of all record holders in Table 2.14a then in Tables 2.14b and 2.14c he has a 50% chance to perform a record linkage attack since both of them are 2-anonymous. In Table 2.14c each record have a generalised range that is either smaller or equal to the corresponding range in the corresponding record and $\mathcal{QI}$ attribute in Table 2.14b. This lead to achieving a better utility using the non-homogenous generalisation, regardless of the information metric used.

k-concealment Tassa et al. [245] proposed the *k-concealment* which was based on the (k,k)-anonymisation that examined previously. The goal in *k-concealment* is the

Age	Gender	Zipcode	Disease
30	M	10152	Viral Infection
28	F	10157	Diabetes
15	M	10118	Cancer
48	M	10500	Heart Disease
20	M	10511	Flu

(a) Original Table

Age	Gender	Zipcode	Disease
15 - 30	*	10***	Viral Infection
15 - 30	*	10***	Diabetes
15 - 30	*	10***	Cancer
20 - 48	M	105**	Heart Disease
20 - 48	M	105**	Flu

(b) 2-anonymous using homogeneous generalisation

Age	Gender	Zipcode	Disease
28 - 30	*	1015*	Viral Infection
15 - 28	*	10***	Diabetes
15 - 30	M	10***	Cancer
20 - 48	M	105**	Heart Disease
20 - 48	M	105**	Flu

(c) 2-anonymous using non-homogeneous generalisation

generalisation of $\mathcal{QI}$ values is done in such way that each record becomes computationally-indistinguishable from k-1 others. As a non-homogeneous generalisation scheme does not require each EC to have identical $\mathcal{QI}$ values.

Example 2.6.4. Consider the Table 2.15 with $\mathcal{QI}s$ Age and Zip Code and SA Disease. Table 2.16a corresponds to 2-Anonymised version of Table 2.15, where there are two EC with two and three records.

The Table 2.15 correspond to the original dataset while the Table 2.16a to the 2-Anonymous table of the original. The Table 2.16b corresponds to 2-concealment version of Table 2.15. If the adversary knows all the $\mathcal{QI}s$ of the records still he is unable to link a specific record to less than two records. Assuming that his target is Alice and knows her $\mathcal{QI}s$ values of Age and Zip Code then he cannot conclude which one of the two records, the first and the third, belongs to Alice. Some may argue that the first record is more likely to belongs to Alice, but the authors [245] claims that it is computationally hard to do so.

n**-Confusion** Another relaxation of k-anonymity which is also similar to k-concealment is the n**-Confusion** [230]. n-Confusion provides an equivalent level of privacy as k-anonymity by make the records indistinguishable with respect to the re-identification process. The re-identification process is considered to be a function that given a collection of records from an anonymous table and with any additional information from a space of available auxiliary information returns the probability that are entries from the original table.

Name	Age	Zip Code	Disease
Alice	30	10055	Measles
Bob	21	10055	Flu
Carol	21	10023	Angina
David	55	10165	Flu
Eve	47	10224	Diabetes

Table 2.15: Original Table

Age	Zip Code	Disease
21-30	100**	Measles
21-30	100**	Flu
21-30	100**	Angina
47-55	10***	Flu
47-55	10***	Diabetes

(a) 2-anonymous

Age	Zip Code	Disease
21-30	**10055**	Measles
21	100**	Flu
21-30	100**	Angina
47-55	100**	Flu
47-55	100**	Diabetes

(b) 2-concealment

Table 2.16: 2-anonymous vs 2-concealment [245]

2.6.2.4 MultiRelational k-Anonymity

One assumption made by most approaches to k-anonymity is that each individual has a record stored as one row in a table of a database. The information about an individual can be spread in multiple tables on a database scheme. The protection on record level of k-anonymity does not guarantee protection at record owners level. Even when multiple tables of a database transformed to a single table the protection is insufficient as shown by Negriz et al. [173]. Therefore, Nergiz et al. proposed [173] the *Multirelational k-Anonymity*. The MultiRelational k-Anonymity requires that any record holder, after a join to a person-specific table with all the other record owners, to have at least $k-1$ record owners having the same $\mathcal{QI}\int$ with him.

2.6.2.5 k^m-anonymity

Terrovitis et al. [246] introduced the notion of k^m-anonymity in order to protect the transactional databases which is also a relaxation of the k-anonymity guarantee. Formally, it is defined as:

Definition 2 (k^m-anonymity)**.** A table T is k^m-anonymous if any adversary with background knowledge of up to m items of a transaction $t \in T$, cannot use these items to identify less than k records from T.

A difference of k-anonymity is that there is no distinction between $\mathcal{QI}\int$ and SA on k^m-anonymity. Any item can be sensitive and also any set of items can be used by the adversary to invade the privacy of individuals. It is worth noting that queries with zero answers are also secure since the background knowledge of the adversary cannot be associated with any transaction.

Example 2.6.5. An adversary, that targeted Alice, knows that she has purchased beer, milk and diapers from a store. This kind of knowledge is not hard to obtain. For example, the adversary could see the top of the shopping bag of Alice or could see these items in a photograph that Alice upload in a social network. In the released transactional data the adversary finds the records with these 3 items (beer, milk and diapers). Exploiting this could lead the adversary to limit Aliceś possible transactions to a small set or even uniquely associate her to one transaction. When the adversary knows the entire transaction, it may include sensitive items such as various prescription drugs, breaching Alice's privacy. Assuming that the transactional dataset was anonymised using 5^3-anonymity, this means for these 3 items there would be at least 5 transactions, including Alice's, that would also containing beer, milk and diapers.

2.6.3 Countermeasures to Attribute Linkage & Table Linkage

To counter the Attribute Linkage attacks several notions have been introduced in the literature. The most prominent ones are perhaps ℓ-diversity, t-closeness and β-likeness which are analysed in the next paragraphs. Nonetheless, we discuss other approaches as well and provide specific scenarios in which they can be applied.

2.6.3.1 Confidence bounding

The confidence bounding of adversary confidence of inferring a SA value from a set of $\mathcal{QI}s$ was proposed by Wang et al [266, 265]. In this method, privacy templates are specified by defining which SA value s to protect with a threshold h of the giving $\mathcal{QI}s$. If for the given $\mathcal{QI}$s the confidence of inferring the SA value s is less than h then the privacy template on the table is satisfied. The main advantage of the confidence bounding is that different templates can be set by the data publisher for different values of the SA rather than a global protection policy on all SA values.

2.6.3.2 p-sensitive k-anonymity

Traian Marius Truta and Bindu Vinay [253] proposed the (p)-sensitive k-anonymity.

Definition 3 ((p)-sensitive k-anonymity)**.** An anonymised table T^* satisfy the (p)-sensitive k-anonymity property if it satisfy the k-anonymity, and for each EC in T^*, the number of distinct values for each SA is at least p within the same EC.

ID	**Age**	**Country**	**Zip Code**	**Disease**
1	37	Brazil	24248	HIV
2	38	Mexico	24207	HIV
3	36	Brazil	24206	Cancer
4	35	Mexico	24249	Cancer
5	51	Italy	23053	Diabetes
6	58	Spain	23074	Pneumonia
7	55	Germany	23064	Bronchitis
8	52	Germany	23062	Gastritis
9	43	Brazil	24248	Zika fever
10	47	Mexico	24204	Zika fever
11	46	Mexico	24205	Zika fever
12	45	Brazil	24248	Colitis

(a) Original Dataset.

ID	Age	Country	Zip Code	Disease
1	<40	America	242**	HIV
2	<40	America	242**	HIV
3	<40	America	242**	Cancer
4	<40	America	242**	Cancer
5	>50	Europe	230**	Diabetes
6	>50	Europe	230**	Pneumonia
7	>50	Europe	230**	Bronchitis
8	>50	Europe	230**	Gastritis
9	4*	America	242**	Zika fever
10	4*	America	242**	Zika fever
11	4*	America	242**	Zika fever
12	4*	America	242**	Colitis

(b) (2)-sensitive 4-anonymity.

Table 2.17: (p)-sensitive k-anonymity.

The Table 2.17b, which is a (2)-sensitive 4-anonymous dataset, is used to illustrate a drawback of this approach. The records with IDs from 1 to 4, which consist the first EC, have 2 distinct SA values but still an adversary can conclude that his target has a serious and incurable disease which is a privacy breach.

Two extensions of (p)-sensitive k-anonymity has been proposed in [232] to overcome this problem.

(p^+)-sensitive k-anonymity

Definition 4 ((p^+)-sensitive k-anonymity)**.** An anonymised table T^* satisfies (p^+)-sensitive k-anonymity property if it satisfies k-anonymity, and for each EC in T^*, the number of distinct categories for each sensitive attribute is at least p within the same EC.

An example of how the SA values can be grouped in order to achieve the preferred (p^+)-sensitive k-anonymity is illustrated in Table 2.18.

CategoryID	Sensitive Values	Sensitivity
One	Cancer, HIV	Very High
Two	Pneumonia, Diabetes	High
Three	Bronchitis,Gastritis	Medium
Four	Colitis, Zika fever	Low

Table 2.18: Grouping SA values.

Age	Country	ZipCode	Disease	Category
<50	America	2424*	HIV	One
<50	America	2424*	Cancer	One
<50	America	2424*	Zika fever	Four
<50	America	2424*	Colitis	Four
>50	Europe	230**	Diabetes	Two
>50	Europe	230**	Pneumonia	Two
>50	Europe	230**	Bronchitis	Three
>50	Europe	230**	Gastritis	Three
<50	America	2420*	HIV	One
<50	America	2420*	Cancer	One
<50	America	2420*	Zika fever	Four
<50	America	2420*	Zika fever	Four

Table 2.19: (2^+)-sensitive 4-anonymity.

(p, α)-sensitive k-anonymity This model first have to define an ordinal weight for each category, as shown in Table 2.20, which captures the degree that each specific SA value contributes to the EC. The weight of the specific SA value is equal to the weight of the category that the specific value belongs to. The weight of the EC is the sum of each specific weight of SA values that the EC contains.

CategoryID	Sensitive Values	Weight
One	Cancer, HIV	0
Two	Pneumonia, Diabetes	1/3
Three	Bronchitis,Gastritis	2/3
Four	Colitis, Zika fever	1

Table 2.20: Weights of SA values.

Definition 5 ((p, α)-sensitive k-anonymity)**.** An anonymised table T^* satisfies (p, α)-sensitive k-anonymity property if it satisfies k-anonymity, and each EC has at least p distinct SA values with its total weight at least α.

The example of Table 2.21, which is a (3,1)-sensitive 4-anonymous view of Table 2.17a, where there are at least 3 different values in each EC and the least total weight of the EC is 1.

Age	Country	ZipCode	Disease	Weight	Total
<40	America	242**	HIV	0	1
<40	America	242**	HIV	0	
<40	America	242**	Cancer	0	
<40	America	242**	Zika fever	1	
>40	Europe	230**	Diabetes	1/3	2
>40	Europe	230**	Pneumonia	1/3	
>40	Europe	230**	Bronchitis	2/3	
>40	Europe	230**	Gastritis	2/3	
<40	America	24***	Cancer	0	3
<40	America	24***	Zika fever	1	
<40	America	24***	Zika fever	1	
<40	America	24***	Colitis	1	

Table 2.21: $(3, 1)$-sensitive 4-anonymity.

2.6.3.3 ℓ-diversity

To counter Attribute Linkage attacks Machanavajjhala et al. [153] proposed the ℓ-diversity.

The ℓ-diversity model requires that each equivalence class has at least ℓ different attribute values. A more formal definition of ℓ-diversity as given by Machanavajjhala et al.[153] is the following:

Definition 6 (ℓ-diversity)**.** An EC is ℓ-diverse if there are at least ℓ "well-represented" values for the sensitive attribute. A table T is ℓ-diverse if every equivalence class $\in T$ is ℓ-diverse.

The term "well-represented" in the simplest form is to ensure that there are at least ℓ distinct values for the SA in each EC, which is identical to (p)-sensitive k-anonymity property.

When dealing with more than one SA the use Multi-Attribute ℓ-diversity provides the required privacy.

Definition 7 (Multi-Attribute ℓ-diversity)**.** In a table T with quasi-identifiers $Q_1, Q_2, ..., Q_{m_1}$ and sensitive attributes $S_1, S_2, ..., S_{m_2}$. T is ℓ-diverse if for all $i = 1...m$, the table T is ℓ-diverse when S_i is treated as the sole SA and $\{Q_1, Q_2, ..., Q_{m_1}, S_1, ..., S_{i-1}, S_{i+1}, ..., S_{m_2}\}$ is treated as the $\mathcal{QI}$.

2.6.3.4 ℓ^+-diversity

Liu and Wang [145] proposed an extension, the ℓ^+-diversity, which instead of offering a universal protection on all SA values, it sets user-defined privacy thresholds to each SA value to decrease the information loss.

2.6.3.5 (X,Y) - Privacy

An extension which inserts constraints of confidence bounding to *(X,Y) - Anonymity* have been proposed by Wang and Fung [263], the *(X, Y)-Privacy*. *(X, Y)-Privacy* address the problem of having a value on Y that occurs more often than others, which results the probability of inferring the SA value to be greater than $\frac{1}{k}$. To satisfy *(X, Y)-Privacy* each group x on X has to contain at least k records and for each SA value s on Y the confidence to infer s from x is less than h, where h is the value for the confidence bounding.

2.6.3.6 (α, k)-anonymity

In 2006 Wong et al [274] extended the k-anonymity model by proposing the notion (α, k)-anonymity. Their model in order to protect the sensitive information limits the confidence of the implications that can be made from the $\mathcal{QI}$ to a SA value within a threshold α.The extension of k-anonymity relies on the following requirement.

Definition 8 (α - Deassociation Requirement)**.** Given a table T, an attribute set $\mathcal{QI}$ and a SA value s. Let (E, s) be the set of tuples in EC E containing s for SA and α be a user-specified threshold, where $0 < \alpha < 1$. Dataset T is a α -deassociated with respect to attribute set $\mathcal{QI}$ and the sensitive value s if the relative frequency of s in every equivalence class is less that or equal to α . That is $|(E, s)|/|E| \leq \alpha$ for all equivalence classes E

Based on that, (α, k)-anonymity can be defined as follows:

Definition 9 ((α, k)-anonymity)**.** A table T^* is said to be an (α, k)-anonymisation of the table T if it is modified in such way that satisfies both k-anonymity and α-deassociation properties with respect to $\mathcal{QI}$

2.6.3.7 Personalised Privacy

Personalised Privacy is a solution specific for categorical SA with a taxonomy and have been introduced by Xiao and Tao [278]. In this approach, the users can define their desired level of privacy in contrast to apply the same level of privacy protection to all the individuals. To achieve this personalised privacy the users can select their *guarding nodes* which is a node in the SA taxonomy that a user has no problem to reveal. The requirement of this model is to limit the breach probability of any leaf value under the guarding node within a user-defined threshold. The users who are satisfied with less privacy guarantees than others can reduce the information loss of the model while keeping all record holders satisfied.

2.6.3.8 (k, e) and (ϵ, m) Anonymity

Zhang et al. [291] proposed a permutation-based method, the (k, e)-Anonymity which requires that each EC has at least k different sensitive values, while the range of sensitive values in the EC is no less than a threshold e.The ℓ-diversity works only with categorical SA and (k, e)-Anonymity came to fill the need for protecting numerical SAs and also to provide more accurate aggregate queries than the generalisation-based approaches. A major drawback of (k, e)-Anonymity is that it ignores the distribution of sensitive values within the group and is vulnerable to proximity attacks where the adversary infers with high confidence that the numeric sensitive value of an individual victim falls within a short interval.

Example 2.6.6. Assume that in an EC we have 4 records and the salary is the SA. If the adversary knows that his victim is in this EC and the values of salary are $\{1000, 1030, 1050, 4000\}$

The adversary has 25% chance to discover the real salary of his victim but also has 75% probability that his victim's salary is in the interval $[1000 - 1050]$.

To tackle the proximity attacks on numerical SAs, Li et al. [134] extended (k, e)-anonymity by introducing (ϵ,m)-Anonymity. (ϵ,m)-Anonymity requires that for every SA value in an EC, at most $\frac{1}{m}$ of the records in the EC are allowed to have similar SA values. The similarity is controlled by ϵ and can be the absolute difference, $\mid y - x \mid \leq \epsilon$, or a relative one $\mid y - x \mid \leq \epsilon x$.

The Worst Group Protection (WGP) which introduced by Loukides and Shao [149], is another countermeasure for proximity attacks which handles both numerical and categorical attributes. WGP can be applied without generalising SA values and prevents range disclosure while is taking into consideration the background knowledge the adversary may have. WGP measures the probability of disclosing any range in the least protected group of a table, and captures the way SA values form ranges in a group, based on their frequency and similarity

2.6.3.9 t-closeness

The ℓ-diversity was an important step in protecting against attribute linkage attacks but also had some shortcomings. The following example illustrates them by showing not only that ℓ-diversity may be difficult to achieve, but it may not provide sufficient privacy protection as well.

Example 2.6.7. Assume that the original data on table T has only one SA, the test result for a particular rare virus, which is a boolean value, "Positive" (True) or "Negative" (False). Let us assume that T has 10000 records, 99% of which are negative, so only 1% are positive. The degree of sensitivity for each of these 2 values is cleary different. Supposing that the virus is the HIV, that might have a social impact on the patient, the disclosure of the negative result of that test for an individual, would not matter a lot, as 99% of the population in this example has the same result. On the contrary, if someone had a positive result, he would not want this information disclosed. In order to have a distinct 2-diverse table, there can be at most $10000 \times 1\% = 100$ ECs , so the information loss would be large.

Both *Skewness Attack* and *Similarity Attack* can be seen in action in this example.

- **Skewness Attack:** When the overall distribution is skewed, satisfying ℓ-diversity does not prevent attribute disclosure. Consider the above example and assume that one EC has the same number of positive and negative records. So this satisfies distinct 2-diversity, and its variations, entropy 2-diversity, and any recursive (c,2)-diversity [153] requirement that can be imposed. However, this raises a serious privacy risk, since anyone in the EC could be considered to have 50% possibility of being positive, as compared with the 1% of the overall population. Another thorny issue that is raised, in terms of privacy, is when an EC in our example has 49 positive and only 1 negative record, therefore it is 2-diverse. The overall possibility of being positive is 1%, while in this EC it is raised to 98% which is a serious privacy risk.

- **Similarity Attack:** The semantical closeness of the values is not taken into consideration by the ℓ-diversity. Assuming that an adversary finds the EC of his target in a medical anonymous publication and is a 3-diverse. The three different values of this class are *(gastric ulcer, gastritis, stomach cancer)*. The adversary can deduce that his target has a stomach related problem even if he does not know the exact disease of his target.

The t-closeness was proposed by Li et al [137, 138] as a defence to these attacks. The requirement of t-closeness is that the distribution of a SA in any EC to be close to the distribution of the SA in the original table T.

Definition 10 (t-closeness)**.** An EC satisfies the t-closeness requirement if the distance of a sensitive attribute distribution in this class compared to the distribution of that attribute in the whole table is not greater than a threshold t. A table satisfies the t-closeness requirement if all ECs satisfy the t-closeness requirement.

The *Earth Mover Distance (EMD)* [209] function is used by the t-closeness to measure the closeness between two distributions of SA values and requires the closeness to be within t. The EMD function evaluates the dissimilarity between two multi-dimensional distributions in some feature space where a distance measure between single features, which we call ”the ground distance” is given. EMD lifts this distance from individual features to full distributions.Note that, as proved in [142], the complexity of t-closeness for every constant t such that $0 \leq t < 1$, it is NP-hard to find the optimal t-closeness generalisation of T.

SABRE framework which was proposed by Cao et al. [36] is based on t-closeness principle for both categorical and numerical attributes. The authors argue that algorithms for t-closeness that are built on top of k-anonymisation [137, 138] fail in term of efficiency and speed and their experimental evaluation showed that SABRE achieves superior information quality. In their framework, the data are first partitioned into a set of buckets and then forms ECs by selecting the right amount of records, and this is done by taking into consideration the t-closeness requirement, from each bucket. A microaggregation approach to achieve k-anonymous t-closeness datasets is proposed in [224] where the authors evaluate three different microaggregation based algorithms.

2.6.3.10 β-likeness

t-closeness comes also with limitations and weaknesses. Firstly, it lacks the flexibility of specifying different protection levels for different sensitive values. Secondly, EMD function is not suitable for preventing attribute linkage on numerical sensitive attributes. Moreover, enforcing t-closeness would greatly degrade the data utility because it requires the distribution of sensitive values to be the same in all QI groups.

Example 2.6.8. Suppose that a dataset T with two distinct values on the SA, HIV and Flu. If the overall SA distribution between them is $P = (0.4, 0.6)$, and their distribution in an EC is $Q = (0.5, 0.5)$, then $EMD(P, Q) = 0.1$. Still, if their overall distribution is $P' = (0.01, 0.99)$ and their distribution in an EC is $Q' = (0.11, 0.89)$, then $EMD(P', Q') = 0.1$ again. While both cases satisfy 0.1-closeness, the information gain in the latter case is much larger than the former one. The probability of HIV is raised by 25% from 0.4 to 0.5 at the first case, while in the second case it is raised by 1000% from 0.01 to 0.11. In effect, the two cases are not equal regarding privacy protection. Unfortunately, any function, like EMD, that aggregates absolute differences faces the same problem.

The proposed privacy model of β-likeness by Cao and Karras [35] is aiming categorical data.

Definition 11 (basic β-likeness)**.** Given a table T with a sensitive attribute SA_1, let $V = \{v_1, v_2, ..., v_m\}$ and $P = (p_1, p_2, ..., p_m)$ the overall SA_1 distribution in T. An equivalence class G with SA_1 distribution $Q = (q_1, q_2, ..., q_m)$ is said to satisfy basic β-likeness, if and only if $max\{D(p_i, q_i)|p_i \in P, p_i < q_i\} \leq \beta$, where $\beta > 0$ is a threshold.

For an anonymised table T^* from T to satisfy the β-likeness, all equivalence classes $G \subset T^*$ have to comply with β-likeness requirement.

2.6.3.11 δ-presence

δ-presence [171, 172] was proposed in order to protect datasets from membership disclosure. This metric evaluates the risk of identifying an individual in a table based on generalisation of publicly known data.

Definition 12 (δ-presence)**.** Given an external table T_p and a private table T, we say that δ-presence holds for a generalisation T^* of T, with $\delta = (\delta_{min}, \delta_{max})$ if:

$$\delta_{min} \leq P(t \in T|T^*) \leq \delta_{max}, \forall t \in P$$

A dataset is δ-present when the probability of an individual from a publicly known dataset to be contained in the anonymous publications lies between (δ_{min} , δ_{max}) which is a range of acceptable probabilities. The parameters δ_{min} and δ_{max} define the level of trade-off between the utility and the privacy of the anonymised table T^*. An increase on δ_{min} can lead to better privacy protection since more information is hidden. Nonetheless, it also decreases the utility. Likewise, when δ_{max} declines the utility rise as well with the risk to have a privacy breach. The goal for the data publisher is to select the maximal δ_{min} and the minimal δ_{max} value which will provide the necessary guarantees for his application.

The requirement of all available publicly known data to form a table is a drawback to the implementation of δ-presence. Therefore an extension proposed, the c-confident δ-presence [172], to tackle this issue by relaxing the assumption on the availability of a publicly known table to the data publisher. Only a set of attribute distribution functions is assumed to be available to the data publisher, while the adversary has access only to the public data.

2.6.4 Other Countermeasures to Attacks

2.6.4.1 m-invariance

Most anonymous data publication algorithms do not take into consideration possible future re-publications of the dataset T. Changes on the original dataset may happen through the pass of time, such as new insertions and deletions. Even the slightest changes can lead the anonymisation algorithm to generate a different T^* anonymous table. All these anonymous tables can be used by an adversary to infer sensitive information about the record owners by comparing the records as will be shown in subsection2.7.4. When the deletion of records is not an option on the dataset, then a naïve approach is to anonymise the new records separately and add them to the previous anonymous publication. However, when there is only a small amount of new records, the anonymisation process will introduce a severe information loss. Moreover, the new anonymised records may have a different level of generalisation on the QIs than the previous anonymised table. This could be a problem in the analysis of the data.

To tackle the problem of different levels of generalisation, Sweeney [238] proposed as a solution that the latest anonymous dataset to be always as generalised or more than the previous one and never more specialised. The problem that raised from this approach is that each subsequent release will get increasingly distorted. Xiao and Tao to propose the m-invariance [279] as a countermeasure to this problem. The definition of m-invariance has as prerequisites the following definitions

Definition 13 (Historical Union)**.** At time $n \geq 1$, the historical union $U(n)$ contains all the tuples in T at timestamps $1, 2, \ldots, n$, respectively. Formally

$$U(n) = \bigcup_{j=1}^{n} T(j)$$

Definition 14 (Lifespan)**.** Each tuple $t \in U(n)$ is implicitly associated with a lifespan $[x, y]$, where x is the smallest and the y is the largest integer j that t appears in $T(j)$.

Having the Historical Union and Lifespan defined, the m-invariance now can be defined as follows:

Definition 15 (m-invariance)**.** A sequence releases of $T_1, T_2, \ldots, T_p$ is m-invariant if the following properties are met:

- every $\mathcal{QI}$ group in any T_i has at least m records and all records in $\mathcal{QI}$ group have different values on the sensitive attribute
- for any record r with published lifespan $[x, y]$ where $1 \leq x, y \leq p$, $QI_x, \ldots, QI_y$ have the same set of sensitive values where $QI_x, \ldots, QI_y$ are the generalised $\mathcal{QI}$ groups containing r in $T_x, \ldots, T_y$

To simply put, if a record r has included in other anonymous releases $T_x, ..., T_y$, then all the ECs containing the record r in all $T_x, .., T_y$ must have the same set of SA values in order to satisfy the m-invariance requirement. This happens to make sure that intersection of SA values over all such ECs does not reduce the set of SA values. m-invariance is not a choice when the truthfulness at record level is a requirement since the algorithm adds, the minimal required, counterfeit data records.

2.6.4.2 m-confidentiality

m-confidentiality restricting the probability that an adversary can infer from T^*, the association between any individual and a record in T to $\frac{1}{m}$ by taking into account the adversary's background knowledge.

In order to define the m-confidentiality [272], first there is a need to define the *Credibility* and *Minimality principle*.

Definition 16 (Credibility)**.** Let T^* be a published anonymous table which is generated from T. Consider an individual $o \in O$ and a sensitive value set s in the SA. $Credibility(o, s, K_{ad})$ is the probability that an adversary can infer from T^* and background knowledge K_{ad} that o is associated with s.

The background knowledge mentioned here refers to the *Minimality principle*, as defined below.

Definition 17. Minimality principle Assume that an algorithm A is used to produce an anonymous table T^* that satisfies the requirements R. For any EC in T^* there are no specialization of any QI that can result another table $T^{\#}$ which satisfies the requirements R

Definition 18 (m-confidentiality)**.** A table T is said to satisfy m-confidentiality if, for any individual o and any sensitive value set s, $Credibility(o, s, K_{ad})$ does not exceed $\frac{1}{m}$

2.7 Attacks on the Anonymisation Procedure and Continuous Data Publishing

The focus on the previous analyzed methods was on tackling issues that target the anonymised records. Nevertheless, anonymised records are not the only things that an adversary could use to launch his attack. He could exploit the deterministic procedures of the anonymisation algorithm. In this kind of attacks, the anonymisation guarantees could break by using the deterministic nature of the anonymisation algorithm and a posteriori knowledge from the anonymised dataset.

2.7.1 Algorithm Background Knowledge Attack

In most anonymisation scenarios makes the assumption that the adversary has no knowledge about the anonymisation algorithm that was used by the Data Publisher.

Privacy Model	Record Linkage	Attribute Linkage	Table Linkage	Skewness / Similarity Attacks	Attacks on Continuous Data Publishing	Background Knowledge of PPDP algorithms Attacks
k-anonymity [212, 238]	✓					
(X,Y)-Anonymity [263]	✓					
(k,k)-anonymity [92]	✓					
n-Confusion [230]	✓					
MultiRelational k-Anonymity [173]	✓					
k^m-anonymity [246]	✓					
ϵ-Differential Privacy [63]			✓	✓		
k-concealment [245]	✓					
Confidence bounding [266, 265]		✓				
ℓ-diversity [153]	✓	✓				
(X,Y) - Privacy [263]		✓				
(α, k)-anonymity [274]	✓	✓				
Personalized Privacy [278]		✓				
(k, e)-Anonymity [291]		✓				
(ϵ, m)-Anonymity [134]		✓				
t-closeness [137, 138]		✓		✓		
β-likeness [35]		✓		✓		
δ-presence [171, 172]			✓			
m-invariance [279]					✓	
m-confidentiality [272]						✓

Table 2.22: Attacks each Privacy Model prevents.

However, this is not always the case [280, 116, 117, 45]. The knowledge of the algorithm can be exploited as showed in [272]. The *minimality attack* could break the guarantees of anonymisation algorithms based on the *minimality principle*. It is reminded that the *minimality principle* refers to the property of not generalise, suppress, or distort the data more than it is necessary to achieve the desired protection of privacy.

Example 2.7.1. Assume the anonymous Table 2.23b was produced using as input the original Table 2.23a. Moreover, the adversary has access to the auxiliary Table 2.23c and came to his knowledge that the confidence bounding required the for the $\mathcal{QI}s$ Education and Gender is $h = 60\%$. On Table 2.23c the adversary can see 2 records for the values (PhD,Male) of the $\mathcal{QI}s$ and 5 records for the values (MSc,Male). Moreover, the adversary from the Table 2.23b he can observe that a subtree generalisation has been performed on the $\mathcal{QI}$ Education. This happened because the confidence bounding requirement h was violated on the value Cancer. Only the combination of $\mathcal{QI}$ values (PhD,Male) can lead to this violation. The adversary now can conclude that Achilleas and George have Cancer due to the minimality property of the algorithm.

Education	Gender	Disease
PhD	Male	Cancer
MSc	Male	Viral Infection
MSc	Male	Viral Infection
PhD	Male	Cancer
MSc	Male	Viral Infection
MSc	Male	Viral Infection
MSc	Male	Viral Infection

(a) Original Patient Table.

Education	Gender	Disease
Higher	Male	Cancer
Higher	Male	Cancer
Higher	Male	Viral Infection
Higher	Male	Viral Infection
Higher	Male	Viral Infection
Higher	Male	Viral Infection
Higher	Male	Viral Infection

(b) Anonymous Table.

Name	Age	Gender
Achilleas	PhD	Male
Bob	MSc	Male
Costas	MSc	Male
Dimitris	MSc	Male
Edmond	MSc	Male
Freddy	MSc	Male
George	PhD	Male

(c) Table available to the adversary.

Table 2.23: Algorithm background knowledge attack example.

The authors in [45] identify three properties of the anonymisation methods that are vulnerable to minimality attack.

- **Deterministic Behaviour:** The deterministic nature of the anonymisation algorithms allows the adversary, based on the published anonymous data T^*, to use reverse engineering and reason about the possible datasets T^i that could lead to T^*.

- **Asymmetric Group Choices:** An algorithm could be triggered from a violation on a diversity constrain to merge several smaller groups or to split a larger group. The adversary if understand how the algorithm is triggered then he can infer information about the original dataset.

- **Consideration of $\mathcal{QI}s$ and SAs together:** The common goal of most anonymisation techniques is to break the link between the $\mathcal{QI}s$ and SAs in such way that makes difficult for the adversary to restore it with high confidence. Nevertheless, the algorithmś choices of what to group together conditionally could leak information about the original mapping.

2.7.2 Composition Attack

The Data Publishers may use different privacy models and even when they use the same model they could choose different level of the privacy guarantees.Moreover, the same Data Publisher could anonymise the same dataset with different privacy models and different guarantees for multiple data recipients. Based on the above scenarios Ganta et al. [87] introduced the composition attack.

Example 2.7.2. The Tables 2.24a and 2.24b are 4-anonymous and 6-anonymous respectively, depicting patient data from two hypothetical hospitals. Assuming that Alice's employer knows that Alice is 28 years old, the zip code of her residence is 13012 and that visits both hospitals. He can conclude that Alice has AIDS because the value AIDS it is the only one corresponding to Alice $\mathcal{QI}s$ in both Tables.

2.7.3 deFinetti Attack a.k.a Foreground Knowledge Attack

The adversary in this attack [123, 273] needs to know only the $\mathcal{QI}$s of his target. Then he constructs a machine learning model over the anonymous data. An attack of this kind is successful when the extracted information from the correlations between $\mathcal{QI}s$ and SAs in the anonymised data, change the beliefs of the adversary about his target by increasing his confidence level above the guarantees of the privacy model.

Zip code	Age	Nationality	Condition
130**	<30	*	AIDS
130**	<30	*	Heart Disease
130**	<30	*	Viral Infection
130**	<30	*	Viral Infection
130**	>40	*	Cancer
130**	>40	*	Heart Disease
130**	>40	*	Viral Infection
130**	>40	*	Viral Infection
130**	3*	*	Cancer
130**	3*	*	Cancer
130**	3*	*	Cancer
130**	3*	*	Cancer

(a) T_1 is the 4-anonymous published version of Hospital$_1$.

Zip code	Age	Nationality	Condition
130**	<35	*	AIDS
130**	<35	*	Tuberculosis
130**	<35	*	Flu
130**	<35	*	Tuberculosis
130**	<35	*	Cancer
130**	<35	*	Cancer
130**	>35	*	Cancer
130**	>35	*	Cancer
130**	>35	*	Cancer
130**	>35	*	Tuberculosis
130**	>35	*	Viral Infection
130**	>35	*	Viral Infection

(b) T_2 is the 6-anonymous published version of Hospital$_2$.

Table 2.24: Composition Attack.

2.7.4 Attacks on the continuous data publishing

The possession by the adversary of several versions of anonymised data through time open the door to attacks.

A Data Publisher release a k-anonymised version T_1 of the dataset D_1. New data are collected and appended to D_1, forming the D_2. The k-anonymised data of D_2 is denoted as T_2. As the datasets are different this may result the $\mathcal{QI}$s of anonymous datasets T_1 and T_2 to have a different level of generalisation. An example of this case is the Table 2.25a and the continuous publication Table 2.25b. Note that inside the parentheses are the un-generalised values of the $\mathcal{QI}$s. Having access to both T_1 and T_2, an adversary attempts to breach the privacy of a specific individual. It is also reasonable to assume that the adversary knows the $\mathcal{QI}$s of his victim, and also the chronological order of the publications. Moreover, every record owner in T_1 has also a record in T_2, while the opposite is not necessarily true. A new record in T_2 does not have a corresponding record in T_1. Some of the attacks that can be launched are the following:

- **Forward Attack:** The adversary knowing that his target is in table T_1 uses the table T_2 to identify the target's record. Since his target individual is in T_1, there exists a record in T_2, which matches the $\mathcal{QI}$s and SA value. The records that do not match the SA value in T_2, can be excluded from the possibility to belong to the target.

  Example 2.7.3. Assuming that the $\mathcal{QI}s$ values of the adversarýs target are the [Las Vegas, MSc] and the adversary also knows that his target is included in T_1

Locality	Education	Disease
USA (New York)	MSc	Asthma
USA (New York)	MSc	Asthma
USA (New York)	MSc	Asthma
USA (Las Vegas)	MSc	Cancer
USA (Las Vegas)	MSc	Cancer

(a) Continuous data publishing T_1.

Locality	Education	Disease
New York	Higher (MSc)	Asthma
New York	Higher (MSc)	Asthma
New York	Higher (MSc)	Asthma
Las Vegas	Higher (MSc)	Cancer
Las Vegas	Higher (MSc)	Cancer
Las Vegas	Higher (MSc)	Cancer
Las Vegas	Higher (PhD)	Asthma
Las Vegas	Higher (PhD)	Asthma
New York	Higher (PhD)	Cancer
New York	Higher (MSc)	Cancer

(b) Continuous data publishing T_2.

Table 2.25: Continuous data publishing.

of Table 2.25a. In Table T_1 the adversary observes that his target matches all the records. However, if he examines the second Table T_2, Table 2.25b, together with T_1, then he concludes that the first three records on T_1 cannot all belong to his target. If that was the case then the Table T_2 should have at least three [Las Vegas, Higher, Asthma]. This means the adversary can exclude one of any of these three.

- **Cross Attack:** This attack is very similar to the Forward attack in that the attacker knows that his victim is in the table T_1 and tries to identify the victim's record in T_2 using the table T_1.

  Example 2.7.4. The adversarý knowledge is the same as the previous example. There are 5 matching records in T_2 for $\mathcal{QI}s$ values equal to [Las Vegas,Higher]. This means that at least one of the three records equal to [Las Vegas, Higher, Cancer] was not in Table T_1 since in Table T_1 there are only two records with values [USA , MSc, Cancer]. The adversary now can erase one of these records and decrease the number of possible records in T_2 for his target.

- **Backward Attack:** The adversary in the Backward attack have the knowledge that his target record is only in the table T_2, so if a record in T_2 has a corresponding record in table T_1, then it should be excluded as a possibility to belong to the target. By decreasing the records of the targetś EC in T_2, the privacy guarantees are violated.

Example 2.7.5. In this example, the adversary has a new target with $\mathcal{QI}s$ equal to [New York, MSc] and the knowledge that his targetś record is included only in the 2.25b

T_2. Having 5 matching records for [New York, Higher] in T_2 means that at least one of the three records equal to [New York, Higher,Asthma] must have been in T_1 because in T_2 there are only two other records [Las Vegas, Higher, Asthma] that could correspond to T_1 values [USA , MSc , Asthma]. Therefore at least one of these records with values [New York, Higher,Asthma] must be excluded since preexist adversary's target.

In each of the 3 previous examples, at least one matching record was excluded, so the 5-anonymity of the adversary's target is compromised.

This page is intentionally left blank.

Chapter 3

Inference of QIs attack

From our research in the PPDP domain the following results have emerged. First we will analyze the role inference and the importance of the semantics of data and later on we proposed a solution to the problem of k^m-anonymising continuous data without the use of pre-defined data generalisation hierarchies.

3.1 Introduction

Our motivation behind this work was the medical records and their potentials to lead to new discoveries.

While there is a lot of knowledge and a lot of information shared among researchers, it is understood that in order to proceed we need to share even more information. By fusing this information hidden patterns are expected to be detected leading to new discoveries.

The sensitivity of medical records though does not allow jeopardizing, therefore the information before it is shared has to be preprocessed to anonymise the records of individuals. This will allow the post-processing of the information from others, while simultaneously stop adversaries from extracting the identities of individuals.

Anonymisation algorithms have been introduced to obfuscate the published information, increasing the uncertainty of possible attackers to desirable levels.As we described earlier the published information is a corrupted version of the original dataset, several fields might be suppressed, generalised, perturbated, or even added with noise, decreasing this way the utility of the information. Therefore, the balance between anonymity and utility play a central role in picking which methods are going to be applied to each dataset.

The scope of this work is to highlight the significant challenges in anonymising medical records. As we are going to show, there is another type of attack that can be launched using this type of data that we call "**inference of $\mathcal{QI}s$ attack**". The attack stems from the nature of medical records, but it can be applied to other data as well. The attack has been overlooked by current state of the art, mainly because the fields of the tables to be anonymised are considered independent, which is not the case for medical records and other datasets. Additionally, most of the attacks are trying to expose the sensitive attributes through the combinations of $\mathcal{QI}s$. The attack that we introduce follows another way. It exploits the knowledge derived from the sensitive attributes in some records to recover generalised and/or suppressed $\mathcal{QI}s$, or to break the anonymisation guarantees of anatomized or sliced data, increasing this way the re-identification risk. Therefore, sensitive information about other records can be exposed.

3.2 Attacking and Protecting Anonymised Datasets

3.2.1 Attacks on Anonymised Datasets

The original dataset of our running example which has Explicit Identifiers, $\mathcal{QI}s$ and SA is illustrated in Table 3.1.

Table 3.1: Classification of Attribute - Example

EI	QI			SA
Name	**Gender**	**Zip Code**	**Age**	**Disease**
Alice	F	50100	22	Uterine Cancer
Bob	M	50100	45	Flu
Carolain	F	50100	33	Mastitis
Dennis	M	50100	25	Stomach cancer
Ethan	M	50100	56	HIV
Fay	F	50100	65	Coronary heart disease
George	M	50120	34	Hepatitis
Heather	F	50120	18	Obesity
Ian	M	50120	54	Diabetes
John	M	50120	73	Prostate Cancer
Kate	F	50120	88	Alzheimer
Lea	F	50120	14	Juvenile idiopathic arthritis

As already mentioned the k-anonymity is considered a well-known standard in data anonymisation, however, it can only be considered as a baseline as suffers from a serious limitation. While it takes into consideration the $\mathcal{QI}$ attributes, it ignores the

Table 3.2: k=3 lattice 1,0,1

Gender	Zip Code	Age	Disease
*	50100	14-34	Uterine Cancer
*	50100	45-88	Flu
*	50100	14-34	Mastitis
*	50100	14-34	Stomach cancer
*	50100	45-88	HIV
*	50100	45-88	Coronary heart disease
*	50120	14-34	Hepatitis
*	50120	14-34	Obesity
*	50120	45-88	Diabetes
*	50120	45-88	Prostate Cancer
*	50120	45-88	Alzheimer
*	50120	14-34	Juvenile idiopathic arthritis

Table 3.3: k=3 lattice 0,0,2

Gender	Zip Code	Age	Disease
F	50100	*	Uterine Cancer
M	50100	*	Flu
F	50100	*	Mastitis
M	50100	*	Stomach cancer
M	50100	*	HIV
F	50100	*	Coronary heart disease
M	50120	*	Hepatitis
F	50120	*	Obesity
M	50120	*	Diabetes
M	50120	*	Prostate Cancer
F	50120	*	Alzheimer
F	50120	*	Juvenile idiopathic arthritis

SA attributes. This flaw makes k-anonymity method vulnerable to several attacks, such as the already discussed *Attribute Linkage* attack. ℓ-diversity [153] adds a requirement of at least ℓ different values to be in each EC. However, the ℓ-diversity model failed to protect record privacy against the *Skewness Attack*. To overcome the *Skewness Attack*, Li et al. proposed the notion of t-closeness [137]. t-closeness guarantees that the cumulative difference of SA values inside a EC is not more than a given threshold t when compared to the overall dataset. Afterwards, Brickell and Shmatikov proposed the notion of δ-disclosure privacy [27]. A table is called δ-disclosure private if the distribution of the SA values within each $\mathcal{QI}$ class is roughly the same compared with their distribution in the entire table. The δ-disclosure has the advantage to correctly model disclosures when some value of the SA occurs in certain $\mathcal{QI}$ classes, but not in others. ℓ-diversity, t-closeness and δ-disclosure are additional requirements that are based on k-anonymity. Therefore, they can be considered as extensions and not a replacement of the original concept. Nevertheless, as it going to be shown, none of these extensions can prevent the attack that is illustrated in the next section.

3.3 Inference of QIs Attack

Usually on medical datasets the disease of the record owner is the SA attribute. Definitely, all the aforementioned attacks in Section 2.6 are relevant, nevertheless, our attack follows a different information flow. While most of the attacks try to combine to $\mathcal{QI}$s to expose the SA, our attack goes the other way round. At this point, it has to be highlighted that due to the nature of the SA in medical records, diseases are very of often age or gender dependent, therefore, the value of the SA can expose the value of a generalised or suppressed $\mathcal{QI}$ of the anonymised dataset.

Taking advantage of the last remark, we introduce a new attack, called *inference of $\mathcal{QI}s$ attack*, which is based on the SA values and can break the given anonymity guarantees, such as k-anonymity. To clarify the attack, we should understand that the fields of most of the tables are going to be anonymised can be considered independent. For instance, a table might contain the following fields, gender, age and zip code. These three columns are independent in the sense that there is no logical constraint to assume that a man X years old lives in $XVille$, or that a woman X' years old lives in $XVille'$, unless this is a background knowledge. However, as already implied, this is not the case for medical records, something that is illustrated in the following examples.

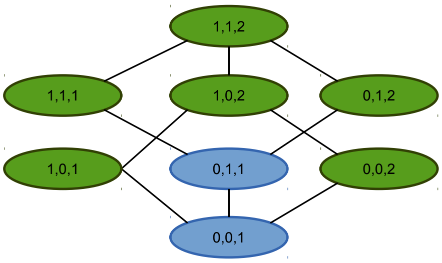


Figure 3.1: Lattice

3.3.1 Examples

Table 3.1 contains the original dataset which can be used to generate a plethora of anonymised tables. Our assumption for the creation of our model is to use 3-anonymity and 3-diversity. All generalised combinations can be shown in Figure 3.1, in which all required guarantees are represented by green nodes, in case they are satisfied, and by blue, in case they are not. According to Table 3.4, it can be understood that in order for the data utility to be maximized, there is a need to select tables with the least information loss. So Tables 3.2 and 3.3 are selected for our attack model, taking into consideration that the $\mathcal{QI}s$ is known by an adversary taken from other databases.

Table 3.4: Possible generalised domains and the regarding information loss. Results produced using Flash [125]

Transf.	Anonymity	Min. Info. Loss	Max. Info. Loss
$[0, 0, 1]$	Not Anonymous	0.0 $[0\%]$	43.0196 $[64, 19\%]$
$[1, 0, 1]$	Anonymous	43.0196 $[64, 19\%]$	43.0196 $[64, 19\%]$
$[0, 0, 2]$	Anonymous	43.0196 $[64, 19\%]$	43.0196 $[64, 19\%]$
$[0, 1, 1]$	Not Anonymous	0.0 $[0\%]$	67.0196 $[100\%]$
$[1, 1, 1]$	Anonymous	43.0196 $[64, 19\%]$	67.0196 $[100\%]$
$[1, 0, 2]$	Anonymous	43.0196 $[64, 19\%]$	67.0196 $[100\%]$
$[0, 1, 2]$	Anonymous	43.0196 $[64, 19\%]$	67.0196 $[100\%]$
$[1, 1, 2]$	Anonymous	67.0196 $[100\%]$	67.0196 $[100\%]$

Example 3.3.1. Table 3.5 is an EC of the original dataset. As it can be seen the gender $\mathcal{QI}$ remained hidden to satisfy the 3-anonymity and the 3-diversity requirement. Despite the anonymity, one can conclude the Gender $\mathcal{QI}$ for two out of three record holders based on specific diseases. For example, uterine cancer and mastitis are female specific diseases. But when it comes to stomach cancer we can foresee that this

record should reflect a male patient. Somebody can lead to this conclusion because of the 3-anonymity and thus not all records can belong to female patients. As a result, Dennis can be re-identified from the original table, since he is the only male that fits in this EC. Lastly, in the other two tuples the EC became from 3-anonymous to 2-anonymous and as such the possibility of a successful linking a tuple to a target has been increased from 33% to 50%.

Table 3.5: example 1 lattice 1,0,1

Gender	Zip Code	Age	Disease
*	50100	14-34	Uterine Cancer
*	50100	14-34	Mastitis
*	50100	14-34	Stomach cancer

Example 3.3.2. When the path (0,0,2) of the lattice is chosen for anonymisation, the Table 3.6 is produced. Here the field Age is suppressed in favour of the anonymisation requirements. In the table above, there are two age-related diseases. In detail, the Alzheimer disease is usually diagnosed in people over 65 years of age and thus somebody can infer that the person suffering from Alzheimer must be Kate. Moreover, the Juvenile idiopathic arthritis refers to a person not greater than the age of 16 and as a result this record must be represented by Lea. By following a similar approach, the last record belongs to Heather. Nevertheless, we should consider that although “obesity” cannot be mapped to a specific age, the adversary might be able to spot a possible candidate if he knows something about the candidate’s appearance.

Table 3.6: example 2 lattice 0,0,2

Gender	Zip Code	Age	Disease
F	50120	*	Obesity
F	50120	*	Alzheimer
F	50120	*	Juvenile idiopathic arthritis

3.3.2 Comparison with similar attacks

The proposed attack differentiates itself from other methods due to the knowledge extraction from the SA value. By doing so, information ($\mathcal{QI}$s) can be inferred for the record holder leading to further exposure.

At this point one can conclude that the attack is a disguised background knowledge attack; however, this is not true as they operate in a completely different way. In our approach, the information flow is from the SA values towards $\mathcal{QI}$s, so it exploits this knowledge to recover the $\mathcal{QI}$s. In contrast, the background knowledge attack, as explained in [153], uses the demographics of $\mathcal{QI}$s to infer the SA.

Similarity attack only has some resemblance to our approach since both are based on the SA values inside an EC. But there is an important difference. Whereas our approach's goal is to gain additional knowledge from the disease to infer for instance the $\mathcal{QI}$ gender, when the latter is generalised for privacy, the Similarity attach groups them and infer the SA of a victim related to a specific type of disease. For instance, uterine or stomach cancer are generalised to the outcome that "victim has cancer".

Moreover, it needs to be stated that the attack described here has no similarity with inference control [76]. In this attack, for the results to be revealed, the user needs to execute multiple database queries, creating "inference channels". As such, an adversary is able to split the query to several others whose intersection can recover sensitive information. This is where our attach method contradicts. It takes into consideration already published and anonymised tables and it concludes by exposing hidden $\mathcal{QI}s$ based on the values of SA.

After all the aforementioned explanations and differences, this proposed attack leads a new way for re-identifying information by using the SA values to expose $\mathcal{QI}s$, reversing the wide adopted attack scenarios, where $\mathcal{QI}s$ are used to exposed SA values. However, the impact of our attack is related to nature of the underlying dataset. The more dependent some $\mathcal{QI}$s are to the SA, the more times such values appear and more data can be linked from the anonymised tables.

3.4 Solutions

Our attack based on the semantics of the records. There is no easy solution that would prevent the attack with generalisation and/or suppression of the records. An obvious way to overcome this is having a domain expert to analyse the output tables and check if there are $\mathcal{QI}s$ that can be inferred from the anonymised table. This may be feasible for a small set of data but unpractical for most of the applications. There is a need for an automated solution that can protect the anonymised medical dataset from correlations that can be used by adversaries to break the privacy guarantees.

In this context, we think that a decision support system could be used paired with the existing algorithms in order to provide an enhanced privacy protection. The system needs a database which contains all the known links between age, gender, ethnicity, location with diseases. Based on that information, the data sets will be preprocessed and all tuples containing such data will be marked. The marked tuples will have to be either suppressed, generalised on another QI , or the linked values will have to be simultaneously generalised to a predefined value in the database. Which action will be taken is depending on the impact of each on information loss. That ensures that the balance between utility and privacy will be kept in a good level for both data recipients and record owners.

3.5 The Inference attack beyond the Generalisation/-Suppression data transformation

There are more techniques, apart from the Generalisation and Suppression, that can be used to anonymise datasets. We will analyse these techniques and how they are related to the inference of $\mathcal{QI}$ attack.

3.5.1 Bucketisation

As we previously described Bucketisation [277] de-associates the relationship between the $\mathcal{QI}$ and the SA, without altering them. Compared to generalisation the anatomised tables give a more accurate answer to aggregate queries that involve $\mathcal{QI}$ values.

Our attack affects Bucketisation because the groups that are formed can be further split into smaller groups i.e based on the age-related diseases.

Example 3.5.1. The Tables 3.7 and 3.8 is an example of bucketsation of the original Table 3.1. As mentioned before this approach has less information loss than k-anonymity since none of the $\mathcal{QI}$s are generalised or suppressed. Moreover, it offers the same privacy guarantee as 3-anonymity, since any tuple from the group 1 has 33% probability of a successful link to its corresponding SA value.

We show that in the group 4 the privacy guarantee is violated when we use the Inference attack. The age-related diseases, Alzheimer and Juvenile idiopathic arthritis, are used by the adversary to link the tuples (F,50120,88) and (F,50120,14) respectively. Obesity can be linked to the one tuple that is left. A complete re-identification of group 4 was performed by our attack.

Table 3.7: Bucketisation - $\mathcal{QI}$ Table

GroupID	**Gender**	**Zip Code**	**Age**
1	F	50100	22
1	F	50100	33
1	F	50100	65
2	M	50100	45
2	M	50100	25
2	M	50100	56
3	M	50120	34
3	M	50120	54
3	M	50120	73
4	F	50120	18
4	F	50120	88
4	F	50120	14

Table 3.8: Bucketisation - Sensitive Table

GroupID	Disease (sensitive)	Count
1	Uterine Cancer	1
1	Mastitis	1
1	Coronary heart disease	1
2	Flu	1
2	Stomach cancer	1
2	HIV	1
3	Hepatitis	1
3	Diabetes	1
3	Prostate Cancer	1
4	Obesity	1
4	Alzheimer	1
4	Juvenile idiopathic arthritis	1

3.5.2 Permutation

Permutation is a method for numerical SA and it is used to improve aggregate queries on such SA. The data records are divided into groups and then the SA values inside the group are shuffled. Permutation is considered beyond the scope of this work since the SA of medical data is usually categorical.

3.5.3 Perturbation methods

In relation to our attack, *Additive noise* can be performed only in numerical SA and the *Synthetic data generation* generates a completely different dataset than the original. *Synthetic data generation* and *Data swapping* do not keep the truthfulness at the record level, which is a requirement for many applications. Moreover, *Data swapping*, unless it takes into consideration the partial $\mathcal{QI}$ dependencies its possible to generate tuples that are obvious not true such as (F, 50120, 18, Prostate Cancer), crippling the utility of the table.

3.5.4 Slicing

Bucketisation can be considered as a special case of Slicing [140], with only two columns, one that contains all the $\mathcal{QI}s$ and the other only the SA. In slicing, columns can be formed with one or more $\mathcal{QI}s$, SA or both.

Example 3.5.2. In this example, as depicted in Table 3.9, Zipcode and Disease form one column. This is very helpful to the data recipient to analyse better their correlations, as attribute correlations are considered an important utility in data publishing.

Performing our attack on the sliced table, Table 3.9, is easy to link the Juvenile idiopathic arthritis with the tuple (F,14) since it is the only one that fits the “juvenile” term. Alzheimer, another age-related disease, is linked to Kate (F,50120,88), in the original Table 3.1. Kate is the only female, with zip code 50120, old enough to have Alzheimer. As in previous example of bucketasation, the remaining tuple with obesity is linked to Heather as she is the only female left with zip code 50120.

Table 3.9: Slicing - Example

Gender & Age	Zip Code& Disease
(F , 14)	(50100 , Coronary heart disease)
(F , 18)	(50120 , Obesity)
(F , 22)	(50120 , Juvenile idiopathic arthritis)
(F , 33)	(50100 , Mastitis)
(F , 65)	(50120 , Alzheimer)
(F , 88)	(50100 , Uterine Cancer)
(M , 25)	(50100 , Flu)
(M , 34)	(50120 ,Prostate Cancer)
(M , 45)	(50120 , Diabetes)
(M , 54)	(50120 , Hepatitis)
(M , 56)	(50100 , HIV)
(M , 73)	(50100 , Stomach cancer)

The interested reader may refer to [147, 148, 241, 210, 222] for more information regarding anonymisation methods dedicated to medical records.

This page is intentionally left blank.

Chapter 4

k^m Anonymity

4.1 Introduction

An anonymous data publication scenario assumes that the original dataset have too many dimensions to expect an adversary to have complete knowledge of his targetś record. For example, such a dataset is collected by a country's tax office. The dataset have dozen of fields concerning financial information of individuals. The adversary may have a partial knowledge of the victimś financial situation. Consider the Table 4.1 that contains income data for individuals which is a sparse multidimensional dataset, that can be represented as a collection of itemsets. Each number on the table represents a different income source as salary, income from agricultural activities, donations, capital gain, etc. Every record holder usually has income from various different subsets from all possible income sources.

It is safe to assume that is a realistic attack scenario of an adversary to be aware of some types of income and not the complete financial data of his target. In our example Alice may be aware that John's salary per annum is 11,000 and additionally that his capital gains are in the range of 18,000 to 22,000. When the unique identifiers are removed in the anonymous publication of the Table 4.1, Alice may use her partial knowledge to identify John's record in the dataset and after that the rest of his income.

A similar dataset, a set-value data such as the market basket data was first presented in [246] where the goal was to anonymise it, using predefined data generalisation hierarchies. Each record is an itemset and every item is a value from a set-value domain $\mathscr{I}$. The anonymisation of such datasets is done by the k^m-anonymity which guarantee that any adversary who is aware of up to m items of a target record cannot use that knowledge to identify less than k individuals in the dataset was analysed in 2.6.2.5.

Name	*Various income sources (annual)*
John	{11000, 11000, 20000, 40000, 40000}
Mary	{11000, 30500, 40000}
Nick	{11000, 11000, 40000, 40000}
Sandy	{11000}
Mark	{20000}

Table 4.1: Original Tax data

Example 4.1.1. The 2^2-anonymous Table 4.2 is an anonymisation of Table 4.1. Any adversary knowing up to 2 values of a target will be incapable to identify less than 2 records. In this dataset to guarantee that level of privacy, using the predefined data hierarchy of Figure 4.1, all values had to be generalised because values {20,000} and {30,500} were rare. However, the same level of privacy can be achieved in Table 4.3 if the values {20,000} and {30,500} are generalised to the range [20,000-30,500]. As we can observe, fewer values are generalised and a smaller information loss is achieved.

Our goal in this work was to provide a k^m-anonymity guarantee to prevent attacks against identity disclosure without a predefined data hierarchy in order to reduce the information loss of the original model. To achieve this we propose a global-recoding generalisation approach that preserves utility by generalising the least number of values necessary to make every combination m values appear in at least k records of the dataset. The assumptions of limited attacker's knowledge as [246] were also made for this approach.The adversary is assumpted to not have negative knowledge, which is reasonable for sparse multidimensional data, and to know up to m values of a target record.

In our approach, there are two main differences from [246]. Firstly we take into consideration the existence of duplicate values in a record and secondly we handle only continuous values. The latter allows us to perform generalisations without the need of a user-defined data generalisation hierarchy. The former helps in preserving statistics,

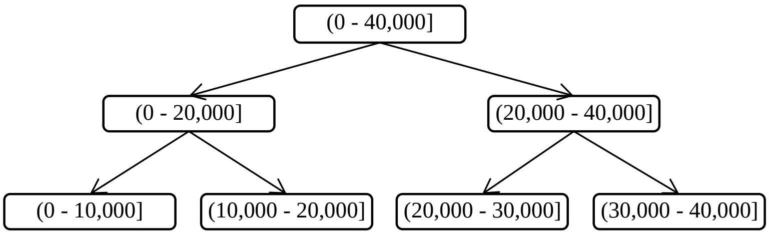


Figure 4.1: Data Generalisation Hierarchy of the data in Table 4.1.

such as the total income of individuals. This means that if two different values in a record are generalised to a common range, they are both kept in the anonymous result, as opposed to [246] which suppresses duplicates. By generalising only the necessary items and to the smallest possible ranges we achieve better information utility, i.e., less information loss.

The main contributions of this work include the following:

- The extension of the problem of anonymising set-valued data [246] to collections of continuous values itemsets;
- Analyse the main differences and challenges of applying k^m-anonymity guarantee to our data scenario;
- Introduce a utility-preserving k^m-anonymisation algorithm for continuous data;
- Evaluation of the algorithm with real-world data and compare the results to the apriori algorithm of [246],

4.2 Problem Definition

Let T be a sparse multidimensional table with continuous attributes $QI_1, QI_2, ..., QI_n$ of the same domain $\mathscr{I}$. Let D be the itemset representation of T where each record is a set of the non-zero values of the respective record of T.

We consider adversaries with limited knowledge to at most m values of a target record t. To identify records on D they use combinations of m values which are rare or unique. If their attack is successful then they can further discover additional information about them, i.e., the remaining values. In the k^m-anonymity there is no distinction between SA and l $\mathcal{QI}s$. Every value is a potential l $\mathcal{QI}$, and all values are equally sensitive as well.

Id	*Various income sources*
1	(10000-20000], (10000-20000], (30000-40000], (30000-40000], (30000-40000]
2	(10000-20000], (30000-40000], (30000-40000]
3	(10000-20000], (10000-20000], (30000-40000], (30000-40000]
4	(10000-20000]
5	(10000-20000]

Table 4.2: 2^2-anonymous table using a data generalisation hierarchy

Id	*Various income sources*
1	11000, 11000, [20000-30500], 40000, 40000
2	11000, [20000-30500], 40000
3	11000, 11000, 40000, 40000
4	11000
5	[20000-30500]

Table 4.3: Continuous 2^2-anonymous table

As stated previously the adversary does not have the knowledge of the zero values because that would correspond to negative knowledge which is more difficult to obtain. For example, it is reasonable for an adversary to know that John's capital gains is 11,000 because he saw a bank statement than to be sure that he doesn't have any capital gains.

Our proposed new algorithm satisfies the k^m-anonymity [246] and ensuring reduced information loss for continuous data, without the need of a user-defined hierarchy.

A non k^m-anonymous dataset D, can be transformed to a dataset $D^\star$, by recoding the values so that $D^\star$ satisfies the k^m-*anonymity* guarantee. In order to succeed that, we generalise only those values that are necessary to make every m-sized combination appear in at least k records, as in Table 4.3. In this case, the generalisation is a set of rules in the form of $v \rightarrow [a, b]$, which map a value v of the original data to a range that includes it.

There can be many different anonymisations of a dataset that satisfy k^m-anonymity for a given adversaryś knowledge limit m, as shown in Tables 4.3 and 4.2. The worst-case scenario would be to anonymise all values to the maximum domain range $\mathscr{I}$. Such a solution is always possible, but it would introduce the highest information loss and the released data would practically have no utility.

The problem of finding the optimal k^m-anonymisation is to find the set of generalisations that satisfy k^m-anonymity and generate the least information loss.

4.3 Anonymisation algorithm

4.3.1 Solution Space

The set of all possible generalisations is the solution space and the accepted solutions are the those that do not violate the k^m-anonymity property. The problem of finding the optimal multidimensional k-anonymity was proven to be NP-hard [160]. As we already mentioned the dataset can be represented as a sparse multidimensional table and this leads to a much larger solution space than that of k-anonymity. This happens for two reasons; (i) k^m-anonymity does not need to form ECs and (ii) without the use a generalisation hierarchy, the set of possible generalisations is significantly larger. Our approach to deal with the increased complexity is a heuristic solution. We take advantage of the *apriori* principle, and perform global-recoding generalisation on the infrequent values at each step of our algorithm, as we explain below.

4.3.2 Dynamic Count Tree

The apriori principle dictates, with a given frequency threshold, any itemset of size n cannot be frequent if any of its subsets is not frequent.This also means that an itemset of size n which is infrequent can not have a superset of size $n + 1, n + 2$, etc. which is frequent.

The proposed algorithm to exploit the apriori principle uses a tree structure similar to the count tree of [246] which is based on the FP-tree of [101]. Each node corresponds to an original value or a generalised value. Progressively itemsets of size i=1, 2, ..., m are examined. The first step is to create a node for every distinct value on the dataset. These nodes are then added to the first level of the tree. Nodes also hold the *support* of the values. A new level of nodes is added to the count tree at each step i of the algorithm. Any path from the tree root to a leaf is an i-sized combination of values. Similarly to nodes, leaves also holds the *support* of the combination in the dataset.

Definition 19. *(support)* The support of a combination of values in a dataset is the number of records that contain this combination.

The algorithm aims for every i-sized combination of values to have support at least k. The core idea is similar to AA of [246] with the exception that in our approach we do not have a predefined generalisation hierarchy. The use of a hierarchy limits the possible generalisations while the absence of it means that an algorithm has to consider all possible generalisations and that is practically impossible. In order to avoid this, we initially add only the original values and their possible combinations. If a generalisation of a value $v \rightarrow g$ is decided by the algorithm, as we will explain later on, the new generalised value g is used and all the original values that fall in its range are replaced with g in the tree. As result the count tree change dynamically depending on the current generalisation rules, as shown in Algorithm 1. In all the subsequent steps, these values will not be considered in the next tree levels, g will be used instead.

Algorithm 1 Incremental Creation of the Dynamic Count Tree **UpdateDCTree**

```
Require: $D$ {Original Dataset}, $T_{i-1}$ {tree of size $i-1$}, $G$ {current generalisations}
Ensure: $T_i$ is the count tree of height $i$.
for every record $t \in D$ do
    for every value $v \in t$ do
        if $\exists$ generalisation range $g \in G$, such that $v \in g$ then
            replace $v$ with $g$.
    for every combination $cmb_i$ of $i$ values in $t$ do
        find path $p_{i-1}$ that contains ($i$-1)-subset of $cmb_i$ (prefix)
        if the $i^{th}$ value exists as a leaf then
            increase its support by 1.
        else
            add the remaining $i^{th}$ value as a leaf under $p_{i-1}$
return $D^\star$
```

4.3.3 Information Loss

To estimate the loss of utility introduced by the value generalisations we use the Normalized Certainty Penalty (NCP) metric [281]. Let v be a value in original domain $\mathscr{I}$. Then:

$$NCP(v) = \begin{cases} 0, & v \text{ } is \text{ } not \text{ } generalised \\ |g_{max} - g_{min}|/|\mathscr{I}|, & otherwise \end{cases}$$

where $[g_{min}, g_{max}]$ is the range to which v is generalised. Intuitively, the NCP tries to capture the degree of generalisation of each value, by considering the ratio of the total domain that are indistinguishable from it.

The total information loss of an anonymous dataset D^* with $|D^*|$ records, is the average NCP of all its values:

$$NCP(D^*) = \frac{\sum_{t_i \in D^*}\{\sum_{v_{i,j} \in t_i} NCP(v_{i,j})\}}{\sum_{t_i \in D^*}|t_i|}$$

where $v_{i,j}$ is the j^{th} value in the i^{th} record and $|t_i|$ is the size of record the i^{th} record.

4.3.4 Algorithm

Our approach is a heuristic global-recoding generalisation algorithm, namely ACD, with m basic steps as shown in the pseudo-code of Algorithm 2. At each step $i = 1, ..., m$ check for privacy violations of itemset of size i. In order to check all possible i-sized combination of values, we use a count tree which is constructed from the Algorithm 1. Every path from the root to a leaf node corresponds to an itemset with support in the dataset equal to the support in the leaf.

Algorithm 2 k^m-Anonymisation of Continuous Data algorithm **ACD**

```
Require: $D$ {Original Dataset}, $m$ {maximum size of attacker's knowledge},
         $k$ {privacy parameter}, $d$ {NCP threshold}
Ensure: $D^\star$ is a $k^m$-anonymous Dataset.
1:  sort tuples with reference to their $f$ values.
2:  $G = \emptyset$
3:  $T_0 = null$
4:  for $i$ = 1, 2, ..., m do
5:      $T_i = UpdateDCTree(D, T_{i-1}, G)$
6:      for every leaf node $f$ in $T_i$ do
7:          if $support(f) < k$ then
8:              $g = findGeneralisation(T_i, f, k, d)$
9:              add generalisation rule: $G = G \cup \{g\}$.
10:             parse $T_i$ in a breadth-first traversal
11:             if there exist one ore more sibling nodes with values $v_1, ..., v_n \in g$ then
12:                 replace values $v_1, ..., v_n$ with $g$
13:                 merge them into a single node $n$
14:                 update $n$'s support
15: return $D^\star$
```

When an i-itemset has less than k support, that means that this combination of values is rare and therefore vulnerable. To increase the path support and in extend to protect the individuals' privacy, some values have to be generalised. A rare value has to merge with a sibling node to form a range that will include both of these values.

The algorithm is a global-recoding one which means that when a generalisation rule $v \rightarrow [v_{min}, v_{max}]$ is decided for a value v, then all instances of value v such that $v' \in [v_{min}, v_{max}]$ have to be generalised also to the range $[v_{min}, v_{max}]$, as shown in lines 10-14 of our algorithm. Each generalisation, at the steps 1,..,i-1 , are kept in a generalisation rules set G (line 9). That way when the i level of Dynamic Count Tree is constructed can take into consideration this set.

Seeking the generalisation with the least possible information loss in the dataset is described in Algorithm 3. If the support of a leaf node is lower than the desired k, then its siblings are the first candidates for merging. This happens cause they share a common prefix, the path from the root to their common parent, which is an itemset of size i-1, and its support is ensured to be $\geq k$ at the previous step of ACD. That means only the leaf values have to be generalised. The function $range(v_1, v_2)$ in line 10 simply returns the range between two values. When $v_1 < v_2$ then $range(v_1, v_2) = [v_1, v_2]$,in case of $v_1 > v_2$ then $range(v_1, v_2) = [v_2, v_1]$. We have a candidate solution when the combined support of the two paths is $\geq k$.

Algorithm 3 Finds a Generalisation that fixes a rare itemset **findGeneralisation**

```
Require: $T_i$ {Count Tree}, $f$ {leaf of a vulnerable itemset path},
         $k$ {privacy parameter}, $d$ {NCP threshold}
Ensure: generalised path of $f$ will have a support $\geq k$.
 1: $n = f$
 2: $S = \emptyset$
 3: for every $s_j$ sibling of node $n$ do
 4:     $S = S \cup \{s_j\}$ {merge candidates}
 5: for every node $s_j \in S$ do
 6:     if the combined support of $s_j$ and $n$ is $\geq k$ then
 7:         $NCP_j = NCP(\{v_n, v_{s_j} \rightarrow range(v_n, v_{s_j})\})$
 8:         if $n$ is not a leaf then
 9:             for every node $nc$ in the path from $n$ to leaf $f$ do
10:                 $NCP_j = NCP_j + NCP(\{v_{nc}, v_{sc_j} \rightarrow range(v_{nc}, v_{sc_j})\})$ {node $sc_j$ is descendant of $s_j$, and it is at the same level as $nc$.}
11: find $s_j \in S$ such that $NCP_j$ is minimum
12: if $NCP_j < d$ then
13:     $g = range(v_n, v_{s_j})$
14: else
15:     let node $n$ be $f$'s parent
16:     goto 2
17: return $g$
```

For all candidate solutions, the NCP is computed. The algorithm selects the solution with the smaller NCP which also has to be smaller than threshold d. If no solution is found in that step then the algorithm parses the problematic path upwards to the root. Now as candidate generalisations are considered merges on both the leaves and their parent nodes, and so on, as shown in line 15 of Algorithm 3.

The worst-case scenario of the algorithm is when all values have to be generalised to the maximum possible range. This means that the ACD algorithm will always find a k^m-anonymous solution.

4.4 Experimental Evaluation

For the evaluation of the algorithm, we used real datasets from the UCI repository [2]. The experiments performed in a computer system with Intel Core 2 Duo CPU at 2.53GHz with 4GB RAM, running Mac OS X 10.8.5. and programmed in C++.

Algorithms. Our algorithm compared to AA [246] the apriori algorithm for k^m-anonymity. The Experimental results indicate that ACD preserves the data utility better than AA, with a small execution time difference.

Data. The selected dataset was the US Census Data 1990 [3] from UCI data mining repository. A subset of 8 numerical attributes was used. These attributes represent different types of income. From a total of 2,458,285 records remained 1,000,000 records when we excluded the records that had zero values on all of these selected attributes. We restricted the value domain to a range of size $|mathcalI| = 1,000$. The average record size was 2.27 and the maximum record size 8 as shown in in Table 4.4.

Dataset	Records	max|t|	avg|t|	Domain size
census	1,000,000	8	2.27	1,000

Table 4.4: Dataset Properties

Parameters. We assessed our algorithm with respect to various parameters:

- k parameter of anonymity
- The limit on attacker's knowledge m
- The dataset size $|D|$
- NCP threshold d

In each test we alter only one of the parameters while keeping the others unchanged. The default settings are k = 10, m = 2, d = 0.001 and $|D|$ = 100000.

Evaluation Metrics. To evaluate the performance of our algorithm we take under consideration the execution time in seconds and the NCP, as a metric for the information loss of the anonymised dataset.

Information Loss. In Figure 4.2 depicted the behaviour of the algorithms in terms of information loss. As k increases, both algorithms increase sublinearly. It is easily observed that the ACD outperforms AA especially for larger values of k, and it preserves more utility than AA for any k, as expected.

As the adversary knowledge m increases, the NCP increases superlinearly for both algorithms. However, for larger m ACD outperforms AA by an order of three times.

The behaviour of NCP threshold d is shown in the first graph of Figure 4.3. It is not possible to compare AA since this parameter is not present in that algorithm. However, for reference we added the NCP of AA of the default parameter settings (k=10 and m=2). For smaller values of d the ACD algorithm results in remarkably improved utility to the released anonymous dataset. As we moving to larger d values the information loss converges with the information loss of algorithm AA. Note that the maximum value of d is 1.

In the second graph of Figure 4.3, we gradually change the dataset size $|D|$. For this experiment, we created 7 random samples of our dataset of sizes 500000, 100000, 50000, 25000, 10000, 5000, 1000 records. Each dataset is a randomly sampled subset of the previous one. As was expected the information loss decreases as the dataset size gets larger and the ACD performs better than AA in all dataset sizes.

Execution Time. We evaluate the computational cost of our algorithm by measuring the execution time of algorithms as depicted in Figures 4.4 and 4.5. For small values of parameter k the execution time is larger and decreases monotonically as k increases. In every test AA is faster than ACD. However, this is a small price to pay and can be overlooked since ACD preserves the data utility better than AA.

The execution time grows sublinearly for both algorithms as the parameter m is increasing. The counts trees that forming for larger m has more levels since more itemsets with bigger size need to be considered by the algorithms.

In Figure 4.5 we can observe the impact of d on the execution time of algorithms. For very small d, the algorithm ACD is slower but this change and slightly outperforms AA for any $d \geq 0.001$.

On the second subfigure of Figure 4.5 depicted the scalability of the algorithms. In both cases the curve grows linear along with the dataset size $|D|$.

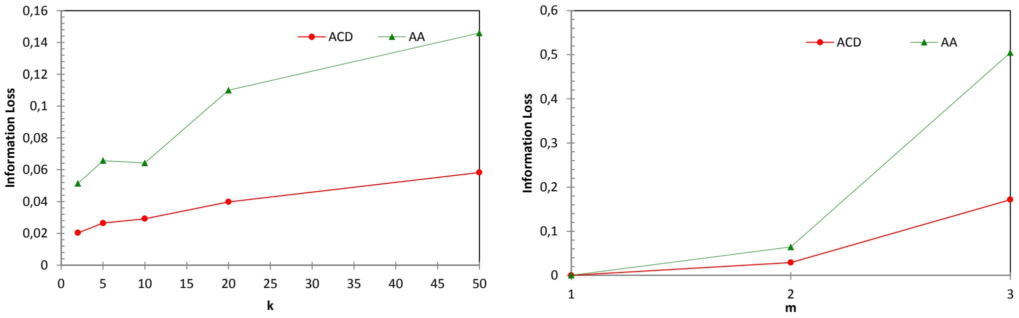


Figure 4.2: Information Loss vs. k and m.

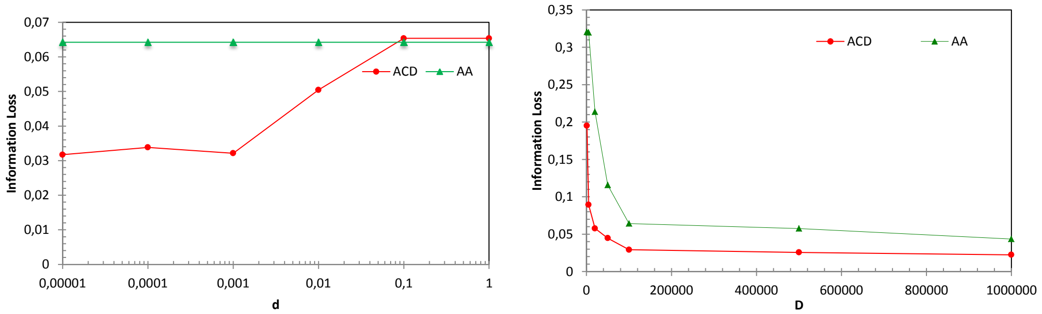


Figure 4.3: Information Loss vs. d and $|D|$.

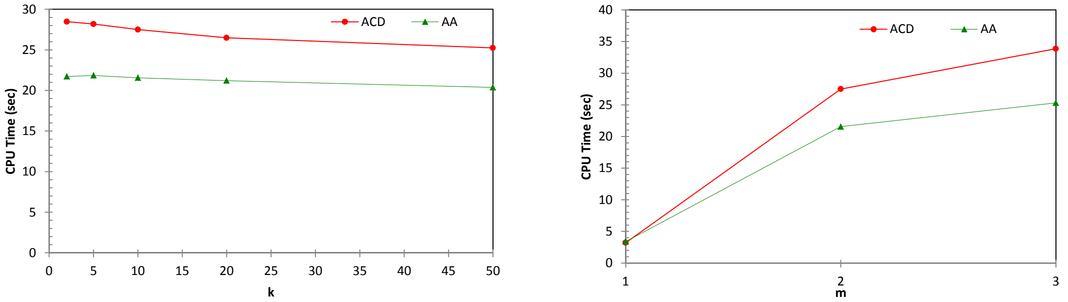


Figure 4.4: Execution Time vs. k and m.

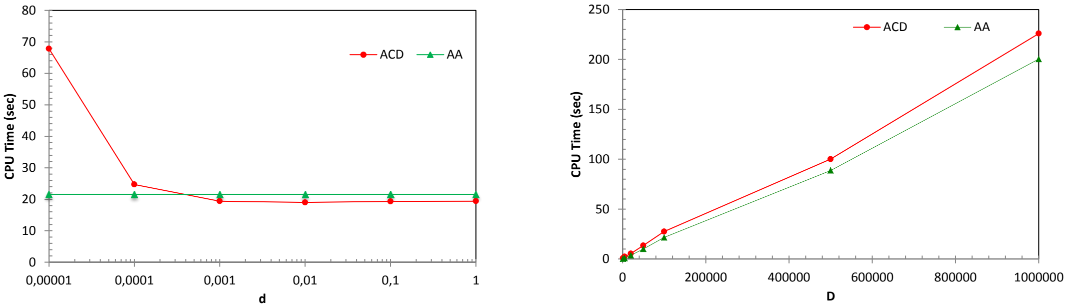


Figure 4.5: Execution Time vs. d and $|D|$.

This page is intentionally left blank.

Part III

Multimedia-based Information Retrieval

This page is intentionally left blank.

Chapter 5

Privacy & Security of Multimedia on Online Social Networks

5.1 Introduction

A few years ago, the preponderance of Online Social Networks (OSNs) was unthinkable, their daily traffic, usage and global acceptance from users, shows that OSNs are here to stay. The digital persona of users has moved from their personal websites to their OSNs profiles. A crucial factor to explain this shift resides in the simplicity that OSNs provide to their users to manage their social lives, increasing their efficiency since they can modify the content of their profiles and control which information about themselves is shared. Furthermore, this information can be edited or even deleted. Hence, users can present themselves as they want, promoting an idealized version of themselves, just like advertisements, which in many cases bare a little resemblance to their actual persona. On the other hand, the recent technological advances in cloud computing and web services, big data storage, cloud computing, semantic web, mobile services, and of course the increased speed of Internet connection have created the necessary requirements to design and develop new services that allow the transmission and exchange of big multimedia files. Blogs, wikis, social networks and publication platforms, have evolved the way that people are connected, creating new communication standards for millions of users around the globe. Social media platforms like Facebook, Google+, Twitter and LinkedIn have completely changed the people's behaviour on the web. Simultaneously, new social media like Pinterest and Instagram highlight that multimedia sharing, more precisely images, either personal or computer generated, are a modern niche market with huge revenues for the service providers, besides, in many cases, events are broadcast in social media before the news appears in common media. Without any doubt, the biggest part of the shared information within social media is multimedia content, uploaded and shared by their users. Nevertheless, the provided security and privacy is often questioned [104, 184, 201, 231, 98, 136, 14]. With regard to security and privacy, malicious users try to exploit software vulnerabilities of the infrastructure to gain access to sensitive personal and professional information. Although this might be very difficult to achieve, attackers may resort to social engineering in order to attack their victims. Tricking users with malicious emails is a very typical approach, not only to steal credentials for OSNs access, but for many other services. It becomes obvious that reposting and republishing of images and multimedia content without any form of ownership, not necessarily about copyrights, can in many cases mislead many users, while on the same time harm the original owner socially and economically. In addition sometimes real users try to bypass privacy measures and disclose information about their peers.

This article exposes and analyzes the security and privacy issues that a user might suffer by sharing content within OSNs and present an up-to-date categorized mapping of these risks, as currently there are many documented issues, few of which are properly addressed. Moreover discusses what are the possible entry points that an attacker will try to use, what he will try to extract and how, what are the possible impacts of his actions and the possible remedies or solutions. It tries to solve the problem of privacy-aware increment for current Online Social Networks. Such increment is possible, even by applying well-known techniques, thus the article focus on how this solution can be achieved and their feasibility within current structures, in terms of implementation effort, processing needs and economic constraints. The main contribution of this work is a new scheme that enables collaboration between OSNs to enhance users' privacy. The novelty of the scheme resides in the fact that it is completely decentralized and does not depend on a trusted third party (TTP). The proposed scheme counters many problems that stem from sharing multimedia content on OSNs such as identity theft, unauthorized content sharing and distortion of malleable content, and also allows a new feature, the shared ownership of multimedia content.

One may argue that the current business model does not allow for such changes as the big "players" do not have the proper incentive to push such solutions forward. They are well established and want to increase their market shares. Therefore, one could claim that cooperation does not seem probable. The recent example of Schema.org[1] exemplifies that this is far from true. The major search engine companies decided to cooperate and create a common framework that helps them to carry out their business

[1] www.schema.org

easier and more efficiently. One should also take into consideration the role of regulatory authorities. The recent deal between EU anti-monopoly authorities and Google[2] signifies that big players can be forced to play with more “open” rules. Thus, developing a common privacy-aware framework for OSNs under the pressure of regulatory authorities[3] is not a far-fetched plan.

It is important to notice that while OSNs disregard each other, there is another link between many of them. Major OSNs may not interact with each other, nevertheless, they allow smaller OSNs to exploit their authentication mechanisms. Therefore, the majority of smaller OSNs are not registering their users directly, but rather obtain user authorization through *e.g.* OAuth[4] to use some of the information from bigger OSNs. This fact indicates that OSNs can further cooperate.

It should be noted that the term of ownership, throughout this work, should not be considered in terms of property or copyright, but it rather refers to the fundamental right to privacy. Users expect that by submitting their personal photos on OSNs, they are free to set their own privacy policies, allowing access only to the users that they decide. The uploaded content is part of their private lives and therefore belongs to them. Therefore, users should be able to selectively reveal themselves to the world [109].

5.2 Related work

5.2.1 Online Social Networks

Online Social Networks (OSNs) could be defined as follows

> *Online Social Networks are web services that provide their users with mechanisms, subject to specific context constraints, to:*
>
> 1. *construct and manage the content and visibility of their profiles within their systems,*
> 2. *define and organize the type of connection with other users,*
> 3. *interact with other users, sharing content and information or even by altering their profiles.*

[2] http://europa.eu/rapid/press-release_IP-14-116_en.htm

[3] https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/Publications/Speeches/2014/14-07-14_PH_for_EV_online_EN.pdf

[4] http://www.oauth.net

This definition identifies the main features of OSNs such as: their dynamic nature, the interaction, the shared content and the context.

The triumph of the OSNs can be attributed to their focus on specific user interests, therefore it is possible to have dating, professional, medical OSNs or OSNs for simple socializing. Moreover, OSNs try to strictly define the type of content that users can share, whether this is multimedia or just text-based information. In this environment, users decide to which users they are related and how, creating the corresponding groups. Depending on their privacy preferences, users can define which information is accessible to which groups of users. Also, users are allowed to interact by exchanging messages and by contributing content to each other's profile, and therefore altering it, by adding, editing and deleting their profiles and shared information according to their desired preferences.

5.2.2 Attacks on OSNs

The wide use of OSNs has piqued the interest of many researchers as well as malicious users. A wide range of attacks has already been documented targeting the users of OSNs. For the sake of completeness, a brief overview of the most important categories of the attacks, which are not related to multimedia content, are presented.

OSNs allows users to communicate, share their thoughts, articles of their interests, suggest movies, books and music and so on. This kind of data can also be used to extract useful information and patterns, which can predict users behaviour and current trends. This knowledge can be used by OSNs to improve their services by offering better personalization strategies, but also by various researchers and third-party applications. To enable the latter, OSNs publish anonymised and aggregated parts of their databases. This way researchers and companies may utilize the given data to find important information. However, the shared data should be anonymous in such a way that no one can infer with great certainty the identity or the attributes of a user [41]. However, as it has been shown, this is not always the case [295]. Currently, there is a lot of effort on anonymising shared data from OSNs in order to develop more efficient and privacy-aware methods of anonymisation [14, 271, 167].

The search capabilities of OSNs have been shown to be vulnerable to crawlers. Automated programs try to reach as many profiles as possible, by utilising the open list of connections that several profiles share. In most cases, crawlers are aimed at the contact information of the users, e.g. email addresses. These emails are active due to the fact that were used to create and activate the OSN accounts. However, in other cases the found contacts are directly used to broadcast spam messages by using the

OSN's infrastructure. Several of these attacks have been documented in the literature [28, 108, 6, 107].

The Group Metamorphosis [48] is another kind of spam. Groups or Pages are communities inside OSNs for users who share specific ideas and/or interests. It has been documented that when the community has a critical mass of users then some administrators may transform the group into a spam platform, posting things beyond the scope of the group.

Social phishing is the evolution of spam attacks. In this attack, a malicious user tries to exploit the access to the victim's personal data such as personal interests or connections. Planning a phishing attack on an individual as shown in [113] has a better click-through rate than typical spamming.

J. Donath, in [61], stated something that we see very often in social networks:

> *"One can have, some claim, as many electronic personas as one has time and energy to create".*

In **Identity Theft** attack, the attacker tries to masquerade as another person to hurt his social profile, or to exploit the trust that other people have in his authority and to obtain money, usually in form of credit. The victim's shared multimedia, which are usually of high quality, can be used to launch attacks in real-life as well, *e.g.* print fake ID cards or company passes. Fraudsters can also extract useful information from the shared multimedia content on OSNs.

In cyberspac,e the replication of victim's account, multimedia content and information, can be achieved easily while this process can even be automated [23]. The following attacks are closely related to the identity theft, and are often regarded as specific cases. If we have replication of the victim's profile in the same OSN, then we have the so-called **Profile Cloning** attack. Otherwise, if the attacker exports the victim's information and multimedia content and creates a profile to another OSN then we have the **Profile Porting** attack. This attack may be more effective for victim impersonation since a search query at an OSN will only return a single profile, the fake one.

Users in the **Sybil Attack** scenario, create multiple accounts to control and affect a result as desired by him and their purpose [62]. It is essentially an escalation of Profile Porting attack. The goal of the adversary can vary from a simple voting scenario to a de-anonymisation attack. A malicious user can also launch an attack on the reputation of a user [103], usually anonymously or/and with the help of a Sybil attack. The attacker spreads, usually false, accusations about the users to draw negative "publicity" that can hurt the victim's social image. Depending on the way in which the victim handles the situation, even if the event proves to be false, the status or the credibility of the victim can be questioned. When a user uploads a multimedia file, setting his desired privacy policy to be shared only with his friends, it implies that he trusts his group of friends, and they will not share or re-upload his file. Nevertheless, in current OSNs his shared multimedia content is usually one click away from bypassing his privacy preferences, leading to the **unauthorized content sharing** attack. Another privacy exposure stems from the use of **static links**, which are used by the majority of OSNs. OSNs use static links to bind the shared content, which can easily be copied and arbitrarily shared on any other medium.

Finally, we find collaborative attacks. OSNs are characterized by the easiness of participation, where the user involvement is very important for the success of the OSN. However, a group of many users can easily abuse this ability and demonstrate a series of coordinate reputation attacks on the content of OSNs, profiles or even whole pages. Collaborative attacks are similar to Sybil attack, just replacing the fake accounts by users with the same goal [252].

5.2.3 Quantifying exposure

Liu and Terzi made the first attempt to quantify the user's privacy content risk [146].

$$PR(j) = \sum_{i=1}^{n} \sum_{k=1}^{l} \beta_{ik} V(i, j, k)$$

where $V(i, j, k)$ is the visibility of user j's value for the attribute i to users who are k hops away from j and β_{ik} is the privacy sensitivity of attribute i.

Domingo-Ferrer, using the above quantification of risk, proposed protocols that assist users in making rational decisions about which attributes should be revealed to other users of an OSN [54]. The decisions are based on the utility that the disclosure of an attribute offers to the rest of the users, so that a correlated equilibrium among users is achieved. Going a step further, Domingo-Ferrer proposed the notion of coprivacy or cooperative privacy, where users cooperate in providing each other with feedback on which attributes to disclose to preserve their privacy [55]. The more someone helps others in preserving their privacy, the more his privacy is preserved. Closely related, but more focused on OSNs, is the approach of Hu et al. [106], which tries to provide a mechanism that addresses to the identification and resolution of privacy conflicts for collaborative data sharing.

Talukder et al. introduced Privometer [243], as a tool to alert users about their exposure over OSNs, this tool is implemented for Facebook and focuses mainly on sex and political view exposure of the users based on their posts.

5.2.4 Tools for privacy in OSNs

It should be noted that very closely related to our research is the work on Social Identity Management (SIdM). Which can be understood as the set of methods that OSNs use to allow users to disclose information to specific groups of their contacts. This allows them to manage the attributes and information that they disclose regarding their social identities/roles, attributed by others or themselves. As it becomes apparent, SIdM is not only focused on multimedia content, but any attribute that an OSN user can have[5]. For more details, the interested reader is referred to [175, 206].

Nowadays, several solutions related to users' privacy on OSNs, have been proposed. The bulk of these solutions are external applications and not native solutions, having several drawbacks that do not allow their wide adoption. Even, many of them are *experimental* solutions or proofs of concept. Therefore, the interface and support are quite limited. The nature of these tools might even bypass the terms of service of each OSN *e.g.* as they use cryptographic or steganographic methods, which hide the main source of income of OSNs, information. Therefore, the solutions which are discussed in the following paragraphs are not widely used and this why users are unaware of their existence. For instance, completely decentralized OSN architectures like Diaspora[6], Safebook [49] and OneSocialWeb[7] never managed to attract massive amounts of users to change the rules of the game.

In NOYB [99], groups of users share a key and break their personal information into "atoms" which are then permuted with the "atoms" of other users, using the key to generate the permutation. Thus, the real information is hidden from the OSN and the users who do not have the key.

Persona [15], allows users to encrypt their data and exchange a public key with selected users. This way, Persona provides an attribute-based encryption to users' data, which allows them to apply their desired privacy policies regarding data access. EASiER [114] extends Persona, by creating decryption keys that are associated with each user, allowing data access, only when a user contacts the proxy with the appropriate key. Another encryption based tool is FlyByNight [150]. It mainly uses public

[5] The shared multimedia content can be considered an abstract attribute of a user's profile.

[6] https://joindiaspora.com

[7] http://onesocialweb.org/about.html

key encryption algorithms to exchange users' messages on Facebook. Scramble [20] is a Firefox extension which allows OSNs users to encrypt their uploaded content storing it either at a TinyLink server or the OSN.

PrivacyJudge [127] allows users to manage who can access their posted content, hosted on their own or a trusted third party privacy server. Data treatment is specified by labels to reduce the risk of accidental exposure of personal information. In this area, there exists Lockr [249] an access control system and Facecloak's [151] which obfuscates users' profiles by providing fake information to the OSN and storing personal information on an application server in encrypted form. Patsakis and Solanas [186] propose a novel cryptography methodology for sharing data within social networks, while users encrypt all their data, they create small encrypted keyword dictionaries on the data that they are willing to share. By sharing the dictionaries' decryption keys with advertising companies, users allow them to mine their data. If they find a promising profile, they can place a bid to access the full data.

In [51], De Christofaro *et al.* propose the use of private set intersection (PSI) protocols to disclose only the common connections that two users have. On the other hand, based on PSI protocols, Li *et al.* introduce a recommender system for social networks, which matches users with similar interests, without disclosing their preferences [135].

Two solutions, X-pire! [13] and unFriendly [248], follows radically different approaches. In the case of X-pire!, users set expiration dates for their shared multimedia content, to make them unavailable after that date. On the other hand, unFriendly proposes a solution for enforcing multi-party privacy in published photos so that they are co-managed by the people who are depicted in them.

PlusPrivacy [8] is a tool which provides a unified dashboard for protecting users from a variety of privacy threats. Regarding multimedia content, this tool applies the most privacy-friendly values automatically across multiple OSNS with a single push of a button. The users do not have to dig into the privacy settings pages of each of their social network accounts.

Facebook [1] and Google+[9] have recently started using face recognition services. The focus of these services is mainly to tag the shared content and allow better search capabilities. However, one could claim that these services could also be used to counter ID theft attacks. The main drawback of these solutions is that the images have to be checked against huge amounts of photos so that even if the identification error is quite

[8] https://plusprivacy.com/

[9] https://support.google.com/plus/answer/2370300?hl=en

small, the total amount of false positives creates an enormous manual processing. On top of that, it should be considered the fact that many users tend to use and share many common pictures, which would issue many false alarms.

5.3 Attack vectors

In this section, we discuss the origins of the possible attacks.

Multimedia: Multimedia files are a rich source of information. It is often said that "a picture is worth a thousand words". Every day numerous of multimedia files are stored on the OSNs. A lot of these files contain sensitive and personal content of the users. Therefore, the multimedia files themselves can be considered a threat to the user if is not treated responsibly.

Malware : Malware intentions are to harm the users or their computer systems. Examples of malware are the keyloggers, ransomware and any other software that can be used in order to exploit any vulnerabilities of the operating system and installed programs.

Misplaced Trust : The lack of any kind of verification of the identity of a user on OSNs lead people to use a naïve approach. In order to trust someone users check their profile picture and the common friends. An adversary with a little effort can effectively attack his target. When the imposter gets in the friend list of the user, he gets access to information and multimedia content that is meant only for trusted users.

Phishing: Phishing is considered a social engineering attack. The adversary is "phishing" usually by setting a legitimate-looking website or email and by pretending that represents a legitimate and credible entity that his victim trusts. By doing so, the adversary attemps to steal the credentials of the victims for the targeted service, for example, e-banking, email or OSN account. If the victim is tricked successfully then his account is compromised.

Hijacking: Profile hijacking is when an adversary breaks into the account of a user and starts to impersonates him to harm his reputation or to run a scam

URL shortening/redirection : URL shortening is a method where a Uniform Resource Locator (URL) may be made substantially shorter and still direct to the address. With the launch of Twitter which limit the length of the messages, the shortened URLs became very popular and common. The real address behind a shortened URL cannot be determined visually or even by looking at the source code, therefore, someone who click it could end up in a legitimate web page, but he could also be led to scams, malicious sites, or other sites that the user did not intend to visit.

Lack of Policies: Unfortunately, OSNs do not have policies to govern every possible privacy issue or to allow fine-grained user customization. Since there is a wide range of possible scenarios of human interaction, a malicious user could take advantage of this. Moreover, as it is going to be discussed later, often several events, such as content re-uploading are not handled by any OSN policy, exposing their users greatly.

Platform vulnerabilities: OSNs as software platforms they do have bugs which an adversary can exploit to bypass users' privacy settings and gain access to personal data[10].

Open access: OSNs are based on the "freemium" model and the registration of their users usually requires only an email authentication. Email providers are also based on'freemium" model. This is a loophole that can be exploited to create multiple and false accounts. It is estimated that between 5.5% and 11.2% of Facebook accounts are fake[11]. Therefore, malicious users can easily launch their attacks anonymously.

5.4 Privacy issues

Privacy is an invaluable human right[12], which in many cases is treated as a product from OSNs, as their mass source of income derives from selling users' preferences to advertising companies. Since this has been documented in the end-user license agreement, it can be considered that users agree to this policy, even if fairer models do exist [186, 187].

E. Houghes in [109] defines privacy as follows:

[10] http://www.neowin.net/news/facebook-photo-exploit-allows-you-to-view-any-albums-of-non-friends

[11] http://investor.fb.com/secfiling.cfm?filingid=1326801-14-7&CIK=1326801

[12] Universal Declaration of Human Rights - Article 12 *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks."*

> *"Privacy is not secrecy. A private matter is something one doesn't want the whole world to know, but a secret matter is something one doesn't want anybody to know. Privacy is the power to selectively reveal oneself to the world."*

However, the ability to fuse information from different sources, even heterogeneous, makes the quest for privacy a rather difficult task in today's interconnected world. OSNs may provide a lot of information about users, using as their source the feedback and interaction of other users. Nevertheless, since users share huge amounts of multimedia in their profiles, a lot of information can be leaked and they can be exposed to great privacy risks without being aware. In an attempt to document the users' privacy exposure due to multimedia sharing we have categorized and analyzed them. A visual representation of these categories is depicted in Figure 5.1.

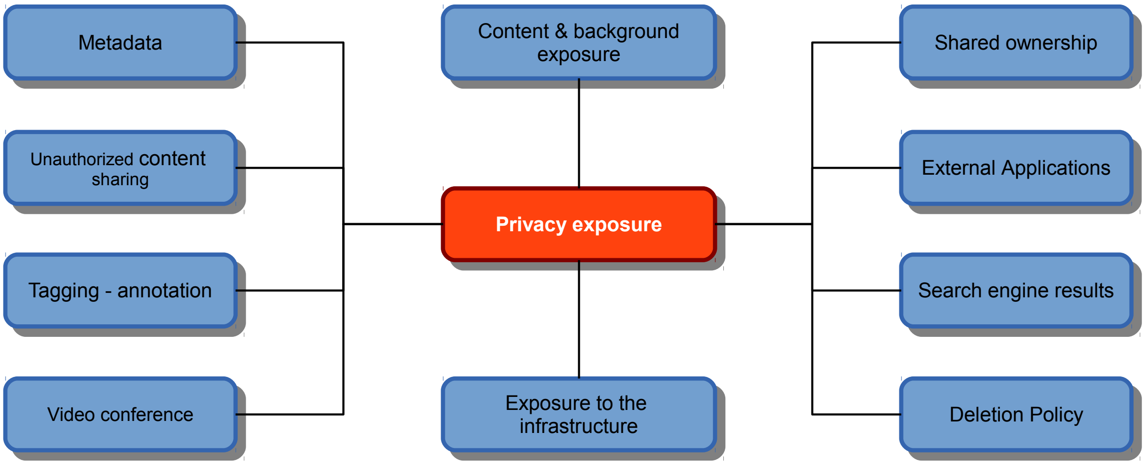


Figure 5.1: Privacy exposure categories.

5.4.1 Content and background exposure

Users are often careful when disclosing textual information over social networks. Therefore, there are very few people sharing their home address or their IDs in OSNs. In contrast, people are not that cautious when it comes to sharing multimedia content, revealing a lot of sensitive information. As an example, users will usually share photographs of their houses, and in many cases their address can be inferred. In other cases, people tweet or post status updates, indicating that they are away from home, e.g. concert, bar, vacations etc, which is more or less indicating that the house is "available" to burglars[13]. An uploaded photo from the current activity can indicate the user location and the duration of his stay, providing additional advantages to the intruders.

In the same context, users have to be aware that burglars may scan the shared pictures for valuable assets. Hence valuable objects depicted in photos or videos can trigger unwanted attention from burglars. Even if the users do not have a direct reference to the date and time of the shared photo or video, several estimations can be made, using background information ranging from sun's location and measuring shadow lengths, or newspapers and people activity.

Also, other forms of privacy exposures can be found, which may include the user or other entities, like a photograph that contains other people. A user may upload this photograph without the consent of others who are present in the photograph and without any notification to them. Depending on the content of the photograph, other users' privacy can be violated or they can be socially discriminated for being caught at the wrong place, at the wrong time. Modern techniques using face and speech recognition can expose many people without their will or any form of notification, i.e. using them in uploaded videos and photographs from public protests, when people share them without anonymising them on their profiles. Furthermore, user profiling can easily be achieved from the shared content and much sensitive information can be deduced as shown in [120].

5.4.2 Metadata

It is possible to define metadata as data about data. They are very useful because they contain additional information about data, so they can be more easily consumed by applications. Especially the multimedia content files contain a lot of information. While this information might be very useful for the user, it might expose him if it is shared.

[13] http://www.pleaserobme.com

A relevant example of such metadata which can expose users is the geolocation tags. Many modern smartphones embed the GPS coordinates in the captured images metadata, which is even more accurate than a street address. This information is rather sensitive as apart from the aforementioned risks, the user's location may disclose many more things about the user e.g. medical condition, political or religious beliefs and so on. Unfortunately, as events have shown, geotagged images may lead even to human casualties[14]. Other image metadata may indicate which camera was used to capture the picture, disclosing its owner and therefore previously unknown connections between users. Depending on the OSN, the metadata are treated differently[15]. Facebook, for example, erases all metadata, while Google+ keeps them considering as sensitive information only the GPS coordinates and prompts users to answer whether they would agree to share them. On the other hand, VKontact[16] by default uses the GPS coordinates to tag the location and uses it to show other users photos from the same location.

5.4.3 Unauthorized content sharing

By sharing content inside an OSN means that a user chooses to disclose this information to a certain group of users. This group can be different for each user's post and is defined by the user's preferences on the privacy settings of the post. Any user that has access to the shared content can bypass the content owner's privacy settings by simple re-uploading the content with different privacy settings this time. This way the content that was intended to be shared among a preselected group of people now has more recipients or even worst to be completely public. Currently this action is not only allowed by the OSNs, but the first user and owner of the content may not become aware of it.

5.4.4 Tagging - Annotation

A lot of OSNs allows the use of tags on the shared multimedia content. Tags can be used for fine-grained search results and interaction among users. Users are allowed to tag photographs and video with any tags they think are appropriate and by doing so they provide additional context information. Nevertheless, this action has some privacy aspects. Some users do not wish to be visually identified, therefore they do

[14] http://www.bbc.co.uk/news/technology-17311702

[15] http://www.embeddedmetadata.org/social-media-test-results.php

[16] http://www.vk.com

not upload any picture of themselves. However, a friend of them can upload such an image and by tagging them to expose their appearance. An extension of the latter is that tagging may allow linking to people, which are not members of any OSN and do not wish to publish any of their information.

5.4.5 Video conference

Many OSNs, like Facebook, have started supporting video conferences. Although this might allow more interaction between users, the problem that arises is that more information can be leaked. Depending on the underlying protocol, the broadcast stream could be intercepted. However, the conference could be easily stored by one of the involved parties to either extort the victim or to manipulate the content and present it accordingly. Besides, possible vulnerabilities in the protocol, or malware could allow the attacker to arbitrarily access the camera and microphone of the victim without notifications.

Experimenting with the latest feature of Facebook to support videoconferences, the author managed to discover another information leak. Since Facebook is using a plugin from Skype to support the video conferences, not all platforms are currently been supported. Therefore, if someone requests a video conference from the other participant, judging on whether the conference can be initiated or not, the use of Windows-based machines can be deduced. While this may be considered minor, it can be escalated afterwards. If the videoconference initializes, then using the log files, each party can see the other's IP address. If their IPs are not spoofed e.g. through proxies, something which is a valid assumption for the vast majority of users, then their location is disclosed with great precision, using off-the-shelf software solutions[17].

5.4.6 Shared ownership

A multimedia file, for several reasons, it is possible to belong to more that one person. The case of two friends who agree to take a photograph together at a social meeting is a good example of shared ownership. The resulting image should belong to both of them. The OSNs do not have any mechanisms for co-ownership. The fact that only one user can set his preferred privacy settings on such content can be considered as a privacy exposure for the co-owner. The OSNs should have a mechanism to set the intersection of co-owners privacy preferences on such content. This would be a more fair solution for all parties.

[17] http://www.visualroute.com/

5.4.7 External Applications

OSNs in order to enrich the user experience and engagement have enabled the development of external applications. These applications have been proven that can be malicious [184] and a security risk. However, other privacy issues are relevant. For example, when users install an application, they have to agree on permissions that the developer would gain on their account.

In the case of Facebook and according to the Facebook's Terms of Service:

> *"When you use an application, the application may ask for your permission to access your content and information as well as content and information that others have shared with you".*

To put it simply, the developer of the application is entitled to use the shared content of the user's friends. This mean that the shared content can be accessed by someone that the user that posted it did not wish, therefore is a clear violation of his privacy. Moreover, the OSNs usually trust all the third-party developers, since there is no restriction on who can develop an application for the OSN platform and no checks if their applications are malicious. The OSN can react on such malicious applications usually through legal actions against the developers related to their Terms of Service agreement.

5.4.8 Search engine results

Search engine crawlers are parsing the OSN that allows them and index the content. By allowing search queries to be executed on OSNs data, informal links between them can be created. Typically, an OSN ignore the existance of other OSNs and treat them as completely different ecosystem. The reality is that many users have profiles in more than one OSN but not on all of them. The search engine results can offer them an insight into what is happening within other OSNs in which the users are not registered. This may seems useful but on the same time opens a backdoor for the users' privacy, because allows the activity of registered users within one OSN to become available not only within one OSN but also to the whole Internet. A user's poor privacy policy could expose him to the whole internet through the search engines.

5.4.9 Deletion Policy

OSNs use the shared content of the users to profile them and to generate income from targeted advertisement. Morever, the users' content is what makes an OSN intersting and people to sign up and visit it. In that sense if a user remove content and information about him, essentially he reduce the income of the OSN. Thus, many OSNs either do not allow users to remove shared content, or allow it with some obstacles in the process (time frames i.e. a photo will not be immediately removed, i.e. users have to pay to remove content[18], etc).

Given that shared content and information in OSNs do not have expiration dates, unified deletion policies raise two critical privacy issues: Are the users entitled to be forgotten? If so, under which conditions? The multimedia content as being the most shared type of content on OSNs is also facing the same privacy issues. Users would like to delete multimedia content from the past e.g. images and videos of themselfs which are not flattering or with previous partners.

5.4.10 Exposure to the infrastructure

It may seem obvious but the exposure of users to the OSN infrastructure might have serious privacy implications. As already mentioned the advertisement is the main source of income on the "freemium" model of most of the OSNs. To raise the efficiency of the advertisement the OSNs are profiling the users and use targeted advertisement. The profiling is based on the information the users provide, with their submitted and shared content in any form, multimedia or not. Due to the disclosure of the role of secret agencies in the Internet[19], the issue becomes even thornier. The service provider has access to users' personal data, which can be very sensitive in the health and medical related OSNs, but disclosures on traditional OSN as Facebook can have serious implications on the users' personal life[20]. Terms of Services of the OSNs can further complicate this issue. For example, Google Plus which is part of the Google services state that:

> *"When you upload or otherwise submit content to our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services,*

[18] http://www.medhelp.org/termsofuse.htm

[19] http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data

[20] http://www.zdnet.com/blog/facebook/facebook-blamed-for-1-in-5-divorces-in-the-us/359

> *and to develop new ones. This license continues even if you stop using our Services (for example, for a business listing you have added to Google Maps). Some Services may offer you ways to access and remove content that has been provided to that Service. Also, in some of our Services, there are terms or settings that narrow the scope of our use of the content submitted in those Services. Make sure you have the necessary rights to grant us this license for any content that you submit to our Services."*

Someone could argue that any of a user submitted multimedia content could be re-published and/or modified without the owner ever knew about it. Even if the user finds out, he would not be able to remove the content and most probably it would be done by non-specifically defined entities, therefore it would be difficult to take any legal actions. Many companies have tried to generalise and simplify the content of their terms of use licenses, nevertheless such wordings can be proved double edged swords, as they open backdoors in users' privacy not only to the service but to malicious employees as well.

5.5 Security issues

There is a plethora of security issues, a great number of which derive from the shared multimedia content. Figure 5.2 shows all the security risks a user is exposed when using multimedia in OSNs.

5.5.1 Unencrypted traffic

Many OSNs use encrypted traffic via SSL only for the login process. This allowed tools such as Firesheep[21], to take advantage of plaintext traffic, by hijacking and intercepting the user sessions. However, there was OSNs which continue operating with unencrypted connections with their users[22],[23]. Using unencrypted connection is a serious problem since the data which is transfered through such connection could be sensitive e.g. from health and medical OSNs. It is reported that such OSNs either still using standard unencrypted http connections or they are using SSL just to send user credentials [202, 214, 180]. As a result, apart from the user credentials, all other content that

[21] http://codebutler.github.io/firesheep/

[22] http://www.motherjones.com/politics/2013/05/shutterfly-teamsnap-eteamz-ssl-hackers-kids-data

[23] The Electronic Frontier Foundation had already warned the Council of Europe for the lack of SSL adoption from OSNs and the impact to the privacy of their users (https://www.eff.org/node/58437).

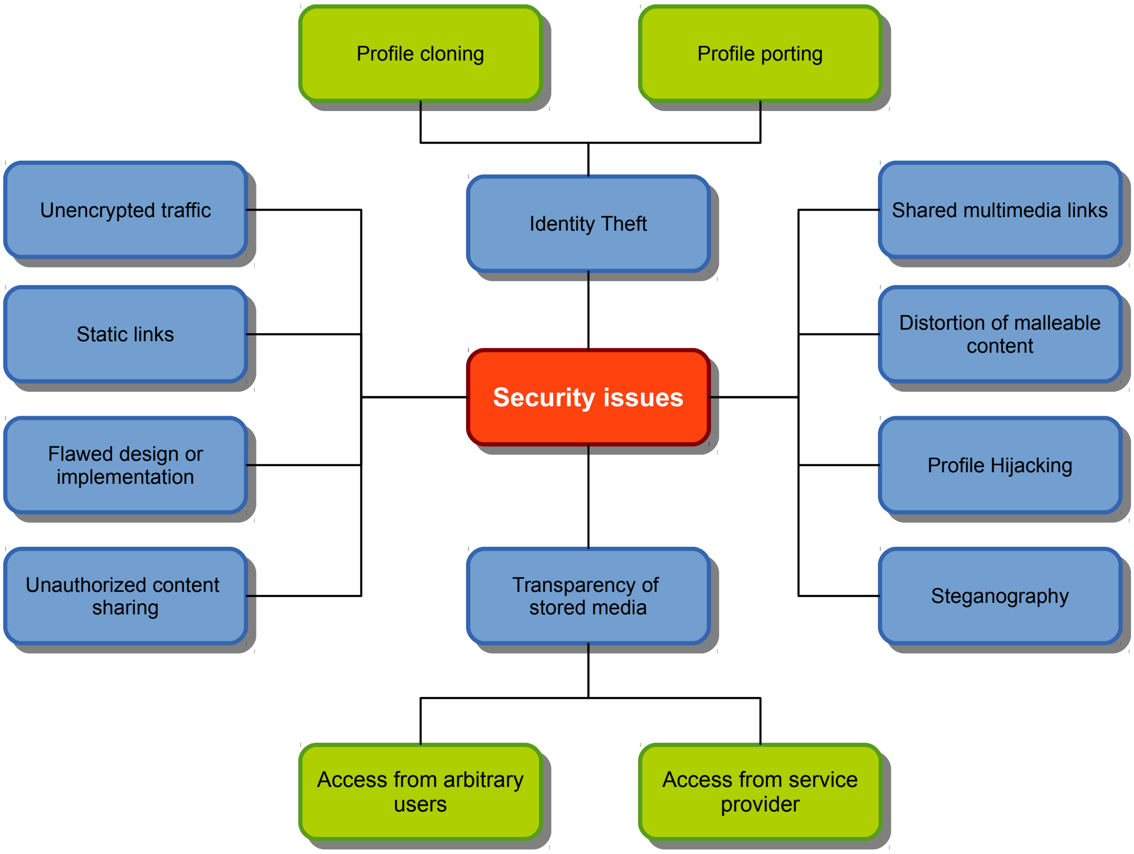

Profile cloning
Profile porting
Unencrypted traffic
Identity Theft
Shared multimedia links
Static links
Distortion of malleable content
Security issues
Flawed design or implementation
Profile Hijacking
Unauthorized content sharing
Transparency of stored media
Steganography
Access from arbitrary users
Access from service provider

Figure 5.2: Security issues.

is being shared e.g. scanned versions of usersḿedical exams expose a huge security vulnerability risk.

5.5.2 Static links

Nowadays almost every OSNs use static links to access multimedia content. Although this is efficient for content distribution, it is definitely not for security and privacy. When OSNs share static links, it gives the ability to users to bypass any privacy and security measures they have. So if a user shares in a restricted group a statically linked image, then every user, having access on it, can share it. Moreover, the link can be shared outside the borders of the OSN. Studies have shown that the use of static links allows people to brute-force them in order to recover other multimedia content [24]. Another interesting threat is the links to a deleted multimedia content. The link usually

[24] http://www.neowin.net/news/facebook-photo-exploit-allows-you-to-view-any-albums-of-non-friends

remains for several days in the OSN after a user requested the deletion. As a result, network administrators, with a simple search in their log files, are able to view the user's browsing history of multimedia content quite easily.

5.5.3 Flawed design/implementation

Since everything human made is expected to have flaws, OSNs have flaws as well. The problem which raised from this flaws is how much can they expose users, how easily can they be exploited and how much effort is needed to trace them. In most of the cases [25] this can be achieved using the shared multimedia content. According to [252] groups of users, real users or bots, can attack the shared multimedia content in order to disable user accounts. The key aspect of such attack is that most of the users on OSNs use nicknames and not their full names. Using nicknames goes against the terms and conditions of the service of several OSNs. As a result, thus their accounts are being blocked or they have to provide more information, such as their real identity, residence etc. to recover their account.

5.5.4 Transparency of stored media

A big issue that is strongly related to static links is the transparency of stored media, which can be understood in two ways. Firstly, the stored multimedia contents are not encrypted, therefore, if someone has a direct link to them they can be accessed without the use of any credentials, bypassing any privacy or security policies set by the user or the OSN. Secondly, there is the transparency towards the service provider. Big OSNs like Facebook or Google+ might have their own data centers, on the other hand, smaller ones do not have this luxury, so they resort to outsourcing their data centers using virtualization or cloud-based technologies. These technologies might reduce scalability and maintenance costs. However, many concerns arise regarding their provided security [26],[27]. In any case, the end user might trust the OSN, but not the cloud service provider which has access to his data[28]. The issue becomes even more thorny due to geospatial and political constrains. Governments and agencies may be granted arbitrary access to foreign citizens' multimedia content without their approval or any kind

[25] http://www.zdnet.com/blog/btl/facebook-acknowledges-photo-privacy-bug-issues-immediate-fix/64819

[26] https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf

[27] http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment

[28] http://slashdot.org/topic/bi/the-windows-flaw-that-cracks-amazon-web-services/

of notification, as the data centers that host this information do not belong to the same country or even continent.

5.5.5 Profile Hijacking

The Profile Hijacking includes attacks where a malicious user tries to take control of another user's profile. The ways this can be achieved is by brute force attacks, phishing, or social engineering. An attacker with the shared multimedia content is possible to have some insights into the user's password. There are some tools like CUPP (Common User Passwords Profiler)[29], which can create a very good dictionary of user's password by providing a proper input. Users should be careful so that the shared multimedia content should not disclose information regarding security questions of user accounts.

5.5.6 Identity Theft

In contrast to the aforementioned attack, there are also attacks in which the attacker is not interested in taking over the user's account. His goal is to create an identical copy of this profile and misleading other users to connect with him. By pretending to be a legitimate credible user he aims to cause reputational damage on the victim or to obtain money by taking advantage of the trust that others have to him. The way to form such attacks can happen by replicating images or other multimedia content from OSNs. Attackers are even able to automate this technique [23]. Moreover, additional information found on OSNs is proven to be extremely helpful for an attacker. For example, a fraudster can extract the age of his "victim" by a birthday picture, make use of the high-quality pictures that can be found on OSNs to print fake ID cards or company passes, etc. There is also a possibility that the attack will take place in the cyberworld and not in the real world as seen above. In this case, the identity theft can be further categorized to profile cloning and profile porting. Such an attack is stealthier, due to the victim's lack of an account on the OSN and as a result, it is more difficult to take precaution measures.

The Identity Fraud Report of 2013[30] by Javelin Strategy & Research indicates that within the U.S.A alone, 12.6 million consumers, 5.26% of U.S adults, were affected in 2012. These attacks enabled fraudsters to steal more than $21 billion in 2012.

[29] http://www.remote-exploit.org/articles/misc_research__amp_code/index.html

[30] https://www.javelinstrategy.com/brochure/276

5.5.7 Distortion of malleable content

Despite the fact that users share a huge amount of multimedia content every day, they are aware that most of this content is malleable. If someone has the intention to harm or make fun of some user, he can achieve it easily with the use of some powerful tools for audio or image processing [234]. With no much of effort, a user's personal photograph, which nowadays can be found in high resolution, can be tampered with to create a new image that looks like real.

5.5.8 Shared multimedia links

Due to the diversity of multimedia formats, it is almost impossible for one framework to support them all. Moreover, since some formats could be eligible for attacks, or their content needs manual check (e.g. interactive flash videos), or their content is embedded in other websites, users are not allowed by the corresponding OSN to share arbitrary multimedia files. For example, someone can share pictures in JPEG or PNG format but not in GIF, which may contain animation and thus it is not that well supported. However, OSNs cannot stop the user's will to share multimedia content outside the OSN. So users are redirected outside the OSN, where they can be without difficulty deceived by malicious entities to install malicious codecs (i.e. using clickjacking techniques), or to visit websites that perform cross site scripting (XSS) attacks. In addition, since the links are static and users are redirected, one could change the content of the initial page, either to maliciously redirect users or to harm the social image of the users that shared it.

5.5.9 Steganography

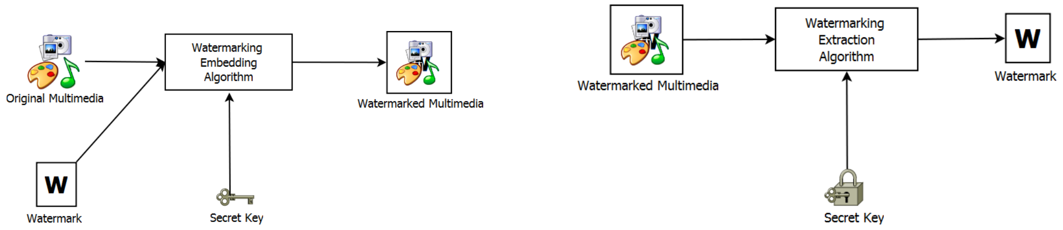
The art of steganography, the technique of concealing information, has been turned into a science occupying a lot of researchers and having many legitimate applications. However, its ability to cover malicious activities can be used within OSNs [38]. Multimedia content can be used as cover objects because of their size. Many authors [256, 176, 242, 164] describe the ability of communication protocols which can be used among users using pictures uploaded to OSNs with embedded information. Although their work is focused on user's privacy, it shows that malicious users can also take advantage of it depending on the embedded information, ranging from terrorist message to child pornography. A problem that can appear from this situation is that users may be blamed for crimes they did not commit e.g. if someone shares a picture he saw from another profile without knowing that it is a cover object. It is clear that

OSNs and users are eventually exposed to such risks, the former by hosting and the latter by sharing multimedia content within OSNs.

5.6 Possible Countermeasures

5.6.1 Watermarking

The goal of steganography is to hide information inside the content that can only be detected by trusted entities. On the other hand, the goal of digital watermarking is to embed information into multimedia in order to prove the origin of the content. Several technics have been proposed to achieve the least possible distortion on the multimedia file e.g. LSB, DCT coefficients etc. While the watermarking is proposed usually as a solution for proving ownership, is also a suitable solution for copy control, fingerprinting and tamper detection. Figure 5.3, depicts the typical structure of a watermarking system. By using a secret key the watermark is scattered in the multimedia content so it is not significantly distorted. Moreover, the watermark cannot be removed and only by having the secret key someone can prove its existence and therefore the origin of multimedia.



(a) Embedding process.

(b) The extraction process.

Figure 5.3: Watermarking.

5.6.2 Encryption of transmitted media

As previously discussed in Section 5.5, many OSNs are not using encrypted traffic or they use it only in the login process. There is no need to argue about the need of using SSL encrypted traffic across all interactions with the service provider. Not using SSL is a serious security risk. In the case of multimedia content the use of SSL guarantee that the uploaded or downloaded content would not be intercepted.

5.6.3 Storage encryption

As discussed in Section 5.4 the multimedia content that users are sharing in many cases can be stored in data centers which are not owned by the OSN and geospatial or political events may expose a lot of users to agencies without their will or any type of notification. The issue is very important given that there are currently many health and medical related OSNs and the shared information is very sensitive. Therefore, whether the user has to be protected from foreign agencies, malicious providers or developers working for the providers, their data should be stored encrypted. There are many cryptographic solutions, mainly based on public key algorithms, which can provide users of OSNs with the required functionality to store and efficiently recover their users' files, without leaking any information to the cloud service provider [261, 287, 182]. Additionally, proxy re-encryption based schemes [10] can guarantee that the users' information will not be leaked within the OSN infrastructure. Another approach, more focused on multimedia, would be the encryption of the multimedia content. While the previous methodology provides arbitrary encryption of data, there exist more focused solutions such as [217]. The advantage of such solutions is that even if someone manages to get a direct link to the shared multimedia content, then the content will not be available unless the user holds the proper decryption key.

5.6.4 Steganalysis

Modern cameras and OSNs enable users to upload high-resolution images, which are large files without raising any suspicions. Nevertheless, they can be used as cover objects to spread malicious content. in those cases, the use of steganalytic software on user multimedia content is considered fundamental. Experiments conducted by the author Show up that such mechanisms do not seem to exist currently in the bulk of major OSNs, or at least their output is not reported to the user. Several OSNs, such as Facebook, may forbid users to use such methods in their terms of service. Nevertheless, they do not seem to block such actions, something that can be exploited. A typical example of the latter is SecretBook[31], a Chrome extension that allows users to exchange secret messages within Facebook, through steganographic methods.

5.6.5 Co-ownership

Models for co-ownership have been proposed by several researchers [225, 227, 226, 106, 179, 159]. OSNs should apply models to allow more than one user to enforce

[31] https://chrome.google.com/webstore/detail/secretbook/plglafijddgpenmohgiemalpcfgjjbph

their privacy policies on the co-owned photos, videos, etc, so that the permissions and restrictions on media are not dictated by the choices of one user and the privacy of all involved users is respected.

5.6.6 Dynamic Links to Content

As already mentioned in Section 5.5, using static links for multimedia content leads to serious privacy risks. OSNs should create dynamic links to the content each time a user request access. The formation of the dynamic links should be subject to the time of request, the IP and MAC address of the user and his credentials. The computation cost of such solution is considered small since it invovles encryption and decryption of small texts. Dynamic links can protect from content exposure within and beyond the OSN platform.

5.6.7 Metadata and background removal

Although many OSNs provide tools to embellish the shared photographs, from simple cropping to applying filters, they do not provide additional functionalities that could help in giving additional privacy to other people. Typical examples are photos from public demonstrations that are uploaded, disclosing the location and political or even religious beliefs of many people. OSNs could provide the functionality for automated detection and removal of faces through e.g. blurring while keeping the necessary information intact. The same functionality could be extended to blurring objects in the background in case the user is interested in hiding some background context.

Additionally, given that not all OSNs follow the same policy towards metadata, all uploaded multimedia files should be stripped of the embedded data unless the user indicates that some of it should be disclosed.

5.6.8 Digital oblivion

In an attempt to offer digital oblivion several solutions have been proposed. Mayer-Schönberger argues that the use of expiration dates is enough to enforce digital forgetting [158]. Moreover, he proposes the implementation of storage devices that can store information with a pre-determined limited lifetime, so that after the lapse of that time frame, the information is automatically deleted. X-pire! [13] aims to allow OSNs' users to store their photos along with an expiration date, after which the images can no longer be accessed. Another approach in which cryptographic primitives are used is proposed in [183]. Using public-key locally-decodable codes the author proposes the gradual decay of the content from a trusted server so that after a certain point in time, or after a certain usage, the content cannot be correctly decrypted and therefore becomes inaccessible.

Domingo-Ferrer proposes a set of protocols where the content creator embeds an expiration date in the content, publishes it and can trace whether someone is using and/or transferring the content after the expiration date [56]. To achieve this, each asset is fingerprinted and the protocols force each entity to cooperate in order to apply the protocol to other assets, as by doing so they know that they are indirectly helping themselves.

In [229], the authors propose the use of a P2P agent community, where the agents negotiate each time which content should be "forgotten" and the content becomes invisible to the users of the OSN.

5.7 Discussion

It has been seen there are many risks involving OSNs and their uses. The Table 5.2, illustrates the risk that has been exposed in this chapter. It is quite clear that the vast amount of possible threats stems from the way that multimedia content is shared within OSNs. In Table 5.3, we illustrate the possible impact that the privacy and security attacks can have on the victim, while Table 5.4 illustrates their difficulty and nature.

It is possible to argue that most of these attacks may be dealt with very well-known solutions. Surely, encryption or digital watermarks, to name two, cannot be considered novelties, nevertheless, the fact that they are not being used as much as they should is, for certain, puzzling. These two solutions, as well as the others, do not come without a cost. The processing cost is quite high, for example, only the cost of using SSL for all transactions reduces the server performance by a factor of around 6 [121, 293]. While this cost is very considerable, the adoption of SSL is a common practice and it is considered to be default nowadays from many webpages and services. Therefore, the fact that it is not fully adopted by all health-related OSNs is unacceptable, as the shared information is very sensitive.

Watermarking and steganalysis of the uploaded content introduce another processing cost, which becomes even bigger if one considers that it has to be applied to all the uploaded multimedia content. Given that most of these services are working

	Watermarking	Encryption of transmitted media	Storage Encryption	Steganalysis	Co-ownership	Dynamic Links to Content	Metadata and background removal	Digital Oblivion
Privacy issues								
Content and Background Exposure							✓	✓
Metadata							✓	
Unauthorized Content Sharing	✓	✓	✓			✓		
Tagging - Annotation					✓			
Video Conference		✓						
Shared Ownership					✓			
External Applications		✓						
Search Engine Results		✓	✓				✓	
Deletion Policy								✓
Exposure to the Infrastructure			✓					
Security issues								
Unencrypted traffic		✓						
Static Links			✓			✓		
Flawed Design / Implementation								
Transparency of Stored Media			✓					
Profile Hijacking		✓						
Identity Theft	✓						✓	
Distortion of malleable content	✓							
Shared Multimedia Links								✓
Steganography				✓				

Table 5.1: Privacy and Security Risks and their Solutions.

under the “freemium” model, a big part of the cost could be reduced either by subscriptions that offer such services as extras, or by elevating the trust to the service, therefore extending their users and customers.

Of course, user awareness is a major issue and users should be warned by OSNs of the exposure that they have and possible threats they might face. Third party solutions might already have been used, nevertheless, their status in terms of acceptance and maturity cannot be considered adequate. OSNs are not expected to provide mechanisms to warn users of what they are about to share will disclose a specific additional information about them, as this is their main source of income. Nevertheless, agents that are developed from third parties could certainly help in this direction, creating a new market and a new line of products.

The Table 5.1 shows how the problems that have been stated could be addressed by existing solutions. As it becomes apparent, the only problem that cannot be tackled by the proposed solutions is the flawed design or implementation. As already discussed, this problem is inherent to almost every software solution, nevertheless, such problems should be quickly resolved when reported and the developers should try to follow common coding standards and principles such as “privacy by design”. Table 5.2 provides an overview of the categorization of the privacy and security issues that are reported in the article. In Table 5.3 we illustrate the impact that each of the reported attacks can have. It is clear that depending on the attacker, the same attack may lead to a completely different impact. Finally, Table 5.4 depicts the difficulty of the attacker to launch an attack on the victim and whether this attack is manual or automated. We should highlight here that the reported difficulty (low, medium, high) is relative to the attacker. For instance, an attack that is based on the exposure of the user to OSN’s infrastructure, cannot be launched by any attacker, but from the OSN itself. In this context, the OSN has to allocate little resources to launch the attack. On the contrary, for an unencrypted traffic attack or for a video conference attack, the attacker is considered to be an average user, which is expected to have limited resources and knowledge. Therefore, the reported difficulty is medium.

	Privacy Related Risk	Security Related Risk
De-Anonymisation Of OSN	✓	
Spam		✓
Social Phishing	✓	✓
Sybil Attack	*	✓
Attacks on Reputation and Trust	*	✓
Collaborative Attack	*	✓
Content and background exposure	✓	
Metadata	✓	
Unauthorized content sharing	✓	
Tagging - annotation	✓	
Video conference	✓	
Shared ownership	✓	
External Applications	✓	✓
Search engine results	✓	
Deletion Policy	✓	
Exposure to the infrastructure	✓	
Unencrypted traffic		✓
Static links		✓
Flawed design/implementation		✓
Transparency of stored media		✓
Profile Hijacking	*	✓
Identity Theft	*	✓
Distortion of malleable content	✓	
Shared multimedia links		✓
Steganography		✓

Table 5.2: Security and privacy issues
* denotes the existence of security/privacy threat with an extension of the attack.

Privacy issues	Information leakage	Location Awareness	Reputation	Account Loss	Ownership Loss	Blackmail Extorsion	Cyberbullying	Cyberstalking
Content and Background Exposure	✓	✓	✓				✓	✓
Metadata	✓	✓						✓
Unauthorized Content Sharing	✓	✓	✓		✓	✓	✓	✓
Tagging - Annotation	✓	✓	✓				✓	✓
Video Conference	✓	✓	✓			✓	✓	
Shared Ownership					✓			
External Applications	✓	✓		✓	✓			
Search Engine Results	✓		✓		✓		✓	✓
Deletion Policy	✓		✓		✓			
Exposure to the Infrastructure	✓				✓			
Security issues								
Unencrypted traffic	✓	✓	✓	✓				✓
Static Links	✓							
Flawed Design / Implementation	✓	✓		✓	✓			✓
Transparency of Stored Media	✓		✓		✓			
Profile Hijacking	✓		✓	✓	✓		✓	
Identity Theft			✓		✓		✓	
Distortion of malleable content			✓				✓	
Shared Multimedia Links			✓	✓				
Steganography			✓					

Table 5.3: Attack impact

Privacy issues	Difficulty	Automated	Manual
Content and Background Exposure	Low		✓
Metadata	Low	✓	
Unauthorized Content Sharing	Low		✓
Tagging - Annotation	Low	✓	
Video Conference	Medium		✓
Shared Ownership	Low		✓
External Applications	High		✓
Search Engine Results	Low	✓	
Deletion Policy	Low	✓	✓
Exposure to the Infrastructure	Low	✓	
Security issues			
Unencrypted traffic	Medium		✓
Static Links	Low	✓	
Flawed Design / Implementation	High		✓
Transparency of Stored Media	Low	✓	
Profile Hijacking	High		✓
Identity Theft	Low		✓
Distortion of malleable content	Low		✓
Shared Multimedia Links	Low	✓	
Steganography	Medium		✓

Table 5.4: Difficulty and nature of attack

Chapter 6

Content management & Watermarking

6.1 Watermarking for Content Management on OSNs

In order to test the possible existence of image watermarking schemes, several experiments were made. The experiments that were conducted in [297] were repeated to test if there is any change in the policies. The original tests were made on the two most widely used OSNs, namely Facebook and Google+. However, we decided to include in our experiments a fast-growing OSN, VK (`vk.com`), which claims to currently host more than 100 million active users. In this experiments were used two groups of images which are going to be referred as Test Set 1 and Test Set 2, using two user accounts, user A and B respectively. The concept was to upload both sets of images on the two accounts and then download again the images from each users' profile and perform some comparisons. Firstly, the images were downloaded from the profile of user A and compared them against their originals. Then, the same procedure was executed for user B. Later, the downloaded images of the two users were compared trying to trace possible differences. The same procedure was repeated for each OSN, from different PCs and at different time frames. These steps allowed us to avoid computer fingerprinting and exclude the time factor from our experiments.

6.2 Experiments

6.2.1 The process

Two groups of images were created. **Test Set 1** includes 40 computer generated and grayscale images from TESTIMAGES[1]. The resolution of 20 of these images is 1200x1200 pixels, while the rest of them have resolution 600x600 pixels. **Test Set 2:** has also 40 images but are closer to what could be characterized as typical user images. This set consists of 20 images with resolution greater than 1200x1200 pixels, which range from 2048x1536 pixels to 3648x2736 pixels. These images were taken from 4 different devices, 7 were taken from the camera of an Apple iPhone 3GS, 6 from a Casio EX-Z1050 camera, 4 from a LG KU 990i mobile and 3 with a Cannon IXUS 130 camera. The rest of the images were taken again from TESTIMAGES, 10 images of 1200x1200 pixels and 10 of 600x600 pixels. The basic image characteristics that are reported in the experimental results were conducted with Matlab.

The process used in the experiments to find out changes in the samples can be seen in Figure 6.1

6.2.2 Results

The results were grouped accordingly. Firstly were presented some general remarks and then were made discussions with the findings for Facebook, Google+, and finally, for VK. For the Test Set 1, the comparison between the downloaded users' images showed that there was no difference in their size or resolution for Google+. The next test was regarding the differences in filesizes of the downloaded images compared to the original ones. In Figure 6.2 we present the histogram regarding the differences in filesizes for Test Set 1. It is obvious that the test set images had no difference compared to the original ones in their filesize when they were uploaded on Google+. However, in almost all of them, it has been noticed a reduction on their filesize, when they were uploaded on Facebook.

Relevant differences were traced in the case of Test Set 2, which consists of high-resolution images. The OSNs have thresholds on the image resolution that can be shared. This is a rectangle of 2048x1536, in portrait or landscape orientation. Beyond this bound, images are resized by both OSNs to fit the optimal resolution within the aforementioned rectangle. In Figure 6.3, we observe again that Google+ does not make any change in the image size if the image is within these bounds. In the Facebook

[1] http://sourceforge.net/projects/test./files/

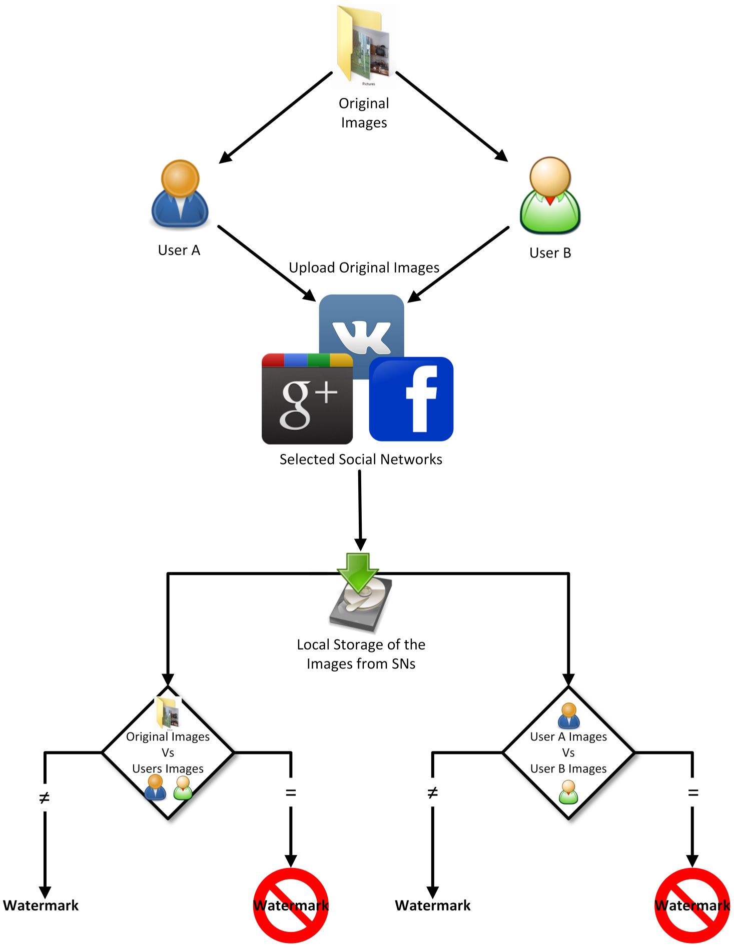


Figure 6.1: The experimental process.

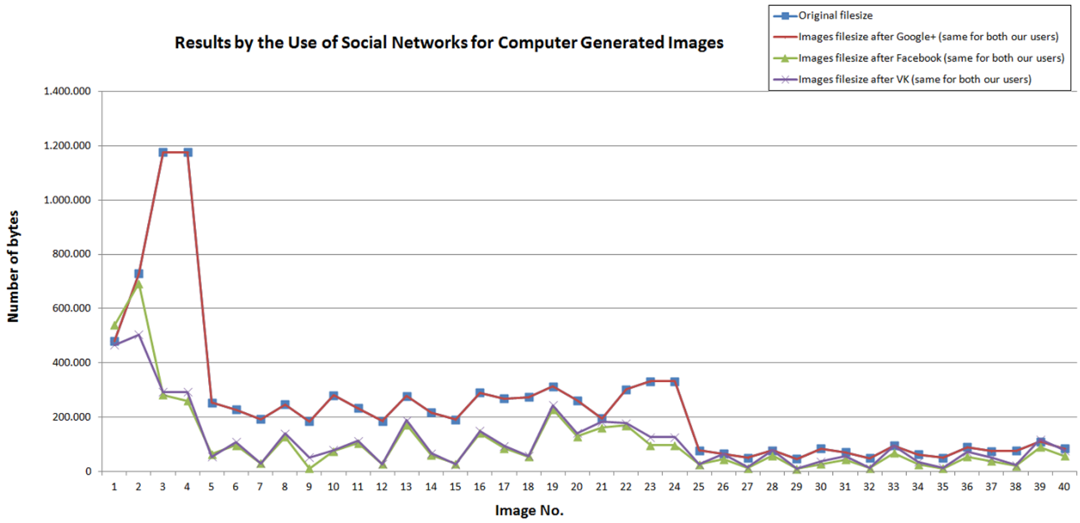


Figure 6.2: Test set 1, image file sizes.

case, however, a big reduction in the filesize is observed, even if the image was of the appropriate resolution. The distortion of several image characteristics is summarized in Table 6.1.

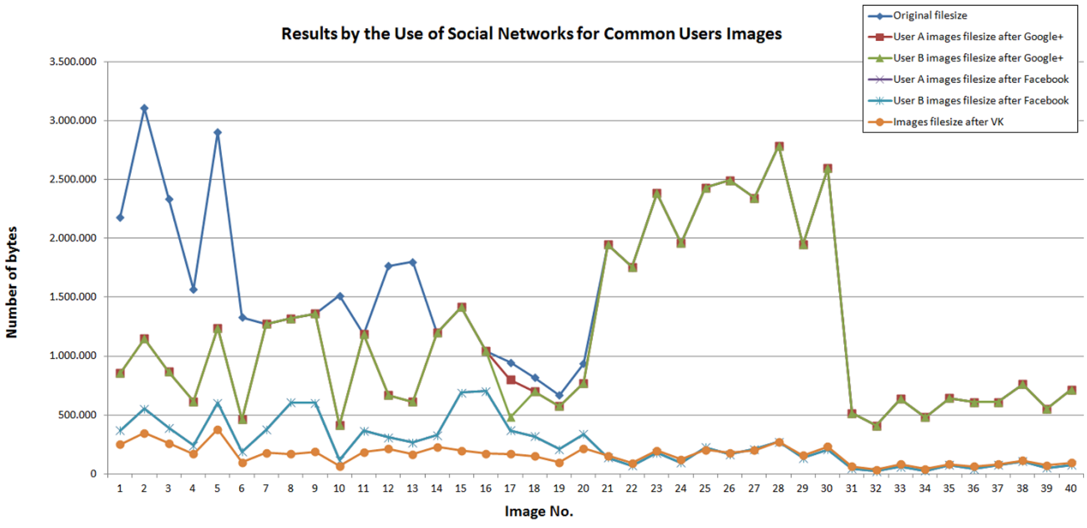
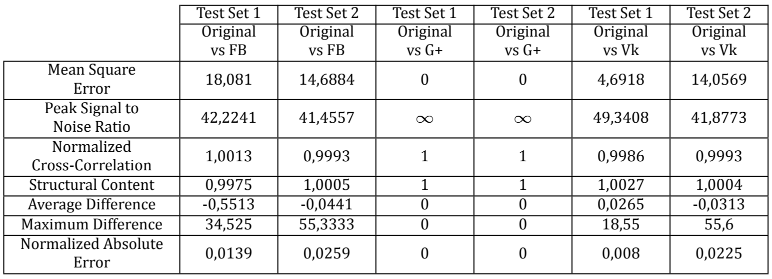


Figure 6.3: Test set 2, image file sizes.

In the case of VK there were more differences. The main difference is that VK has three resolution thresholds for uploaded images, beyond these thresholds, images are resized to fit these boundaries. Therefore, only 30 cases (20 for Test Set 1 and 10 for Test Set 2) fit these boundaries and could be compared against the original ones, all of



	Test Set 1	Test Set 2	Test Set 1	Test Set 2	Test Set 1	Test Set 2
	Original vs FB	Original vs FB	Original vs G+	Original vs G+	Original vs Vk	Original vs Vk
Mean Square Error	18,081	14,6884	0	0	4,6918	14,0569
Peak Signal to Noise Ratio	42,2241	41,4557	∞	∞	49,3408	41,8773
Normalized Cross-Correlation	1,0013	0,9993	1	1	0,9986	0,9993
Structural Content	0,9975	1,0005	1	1	1,0027	1,0004
Average Difference	-0,5513	-0,0441	0	0	0,0265	-0,0313
Maximum Difference	34,525	55,3333	0	0	18,55	55,6
Normalized Absolute Error	0,0139	0,0259	0	0	0,008	0,0225

Table 6.1: Mean values of basic image characteristics,
The table refers to the images that had no change in their resolution.

them being identical. Testing the downloaded images from the profile of user A to the respective from user B, showed again that they are identical, even in the case of size reduction.

6.2.3 Discussion

To facilitate the reader, since there are variations on the results for each OSN, we have kept the discussion of the results for each one of them separate. However, as it will become apparent, with very high probability we can deduce that no watermarking scheme is used by any of them.

6.2.3.1 VK

VK was detected to have three different thresholds depending on the vertical and horizontal ratio. In case of square images that threshold is 1024x1024, for portrait is 768x1024 and for landscape 1280x960. The result shows the complete lack of any watermarking mechanism. The images that are not resized are identical to the original ones. Even when images are resized, both users end up having the exact same images. Therefore, it can be safely deduced that no watermarking has been applied, as this would result to differences in the images of the two users.

6.2.3.2 Google+

For the case of Google+ the results were very similar to the first results in [297], with minor differences. In Google+, when the image resolution does not exceed the aforementioned size threshold, the uploaded image is exactly the same with the original one. Compared to the first experiments, an interesting change was observed in the new images. Whenever Google+ had to resize an image, it inserted an image ID tag in the file's metadata, which was the same for both users. Interestingly, in the new experiments this only happened for one of the photos and for one of the users. The embedded tag cannot in any case be considered as watermark, as it can be removed very easily. Therefore, we can safely deduce in this case that no watermarking is being applied. Moreover, if the image resolution exceeds the threshold, the image is resized, yet, the image is exactly the same for both users. Hence, we may assume that no watermarking is being applied by Google+ on the uploaded images in either case.

6.2.3.3 Facebook

The results for the case of Facebook, indicate that the results in [297] are more or less still valid. Before discussing the findings, it would be necessary to to highlight at this point that Facebook's policy is to convert all uploaded images to the lossy JPEG format. This is blocking users from sharing animated GIFs and distorts lossless formats like PNG. If the images are not resized, then the two downloaded images of the two users are identical. An interesting behaviour was noticed when the images are above the allowed threshold and have to be resized. Specifically, all the images from Test Set 1 do not have any distortions between the users, as their resolution is below the Facebook's thresholds. The images that were different between the two users were separated, and were afterwards checked for steganographic content with stegdetect[2] and stegsecret[3]. The results from both tools were negative, so no steganographic method was traced.

The image differences could hint the existence of an undetected watermarking scheme. However, this approach would be quite peculiar. The distortion is traced only on large photos. Watermarking only high-resolution photos does not sound a good or solid privacy policy, as the allowed thresholds enable attackers to launch attacks on all lower resolution photos. Perhaps, this behaviour could be justified by the existence of a resizing algorithm that uses randomization. A major difference compared to the previous experiments in Facebook, is the image URLs and file names. In the previous experiments, the URL contained the user ID and still does, however, the user ID was embedded in the filename of the downloaded files as well. This enabled third parties to trace the source of an image against others, only from its filename, whenever someone re-posted them or just sent them.

It should be noted that in all three OSNs, the links to media files are static. Users can copy and paste these URLs share them within the same OSN or even worse, share them with people who are not subscribed to the OSN. Moreover, even in the case of

[2] www.outguess.org/detection.php

[3] http://stegsecret.sourceforge.net/

re-uploading an image from another user's profile, there is no notification. This check, in particular, is very easy and lightweight to implement, as it could be checked with the already implemented hashes that are calculated to check the integrity of the uploaded files. Figure 6.3 clearly illustrates that for the images that exceed the resolution threshold, both Google+ and Facebook apply a similar algorithm in terms of compression when images are resized, as the filesizes are almost identical, with the Facebook being a bit more efficient.

6.2.4 Overview of the proposed solution

OSNs can be viewed either as open or closed systems, that have full access to alter the uploaded files. The majority of the users seem not to mind about this kind of distortions, as long as the content is available and without visible distortions to proper correspondents, issued by them. The OSNs' approach, in a conflict of multimedia ownership and misuse, is so far to let users report the offenders. This approach obviously has many disadvantages, as it lets anyone reporting everyone, whether they are the original owners of the content or not. Moreover, a user can report such misuse only when he becomes aware of it by others or by sheer luck. The OSNs currently do not have a sort of policy of notifying users and taking precautional measures about such problems.

As a solution to this problem, it can be proposed a dual watermarking scheme as the first line of defense for both OSNs and users [161, 216]. Of course, user reporting still remains a valuable function in OSNs, but must be used for problems that really require human interference, for example, if someone takes a picture of someone else without his consent, or if the uploaded photo is offensive and misuses the service. As we will see a lot of problems can be solved automatically and with users' notification and awareness.

In this work in order to achieve the objectives, the watermarks should have the following properties.

Invisible The embedded watermarks should be invisible. In contrast to visible watermarking in the invisible watermarking the original multimedia must change in a way that would be imperceptible to the human visual system or the auditory system in case of sound.

Blind An Watermarking algorithm is called blind if does not require access to the original multimedia for detecting and extracting the watermark if the access is needed the algorithm is called not-blind. In a non-blind algorithm for a OSN the space to store the original and the watermarking multimedia can double the needed storage and in case that we decide to save storing space by embedding the watermark on-the-fly, that can have an extreme computational cost. In our approach, we suggest the use of a blind algorithm and the original content to be only in the user's "hands".

Robustness There are three types of watermarks: Fragile, semi-fragile or robust.

- Fragile watermarks are used when the purpose is the complete integrity of the image. Even the slightest modification results to an alert of the watermarking system.
- Semi-Fragile watermarks are used if the goals are only the malicious attacks on the host image and not the common image processing as lossy compression and/or random noise. Any process that has an effect on the content of the image, as cropping or insertion of a new object in the host image, should be noticed by the watermarking system.
- Robust watermarks are used when the final purpose is to prove ownership and that is why they cannot be removed easily and without great degradation of the host image. They must be able to defend against in a wide range of possible attacks.

The interested reader about watermarking and possible attacks, is referred to [47, 46, 268, 196, 195, 258].

The robust watermark it can be used to identify the owner of a media content and may be recovered even if the watermarked media has been processed. Meanwhile, the semi-fragile watermark enables the detection of alterations, malicious or not.

The dual watermarking involves, one robust and one semi-fragile watermarking. In order to clarify the scheme we use the following scenario: User A provides the original multimedia content to the OSN. Then the OSN embeds a robust watermark, which identifies the multimedia content uniquely and relates it to the user A. Afterwards a second semi-fragile watermark is embedded in the content. The dual watermarked content is then stored on the OSN and shared among the users, according to privacy settings set by the user.

The robust watermark identifies the owner of the content so that it can be traced even if the content is tampered. While semi-fragile watermark does not break the first watermark, but at the same time enables the detection of possible alternations from other entities. Figure 6.4 describes the dual watermarking embedding process.

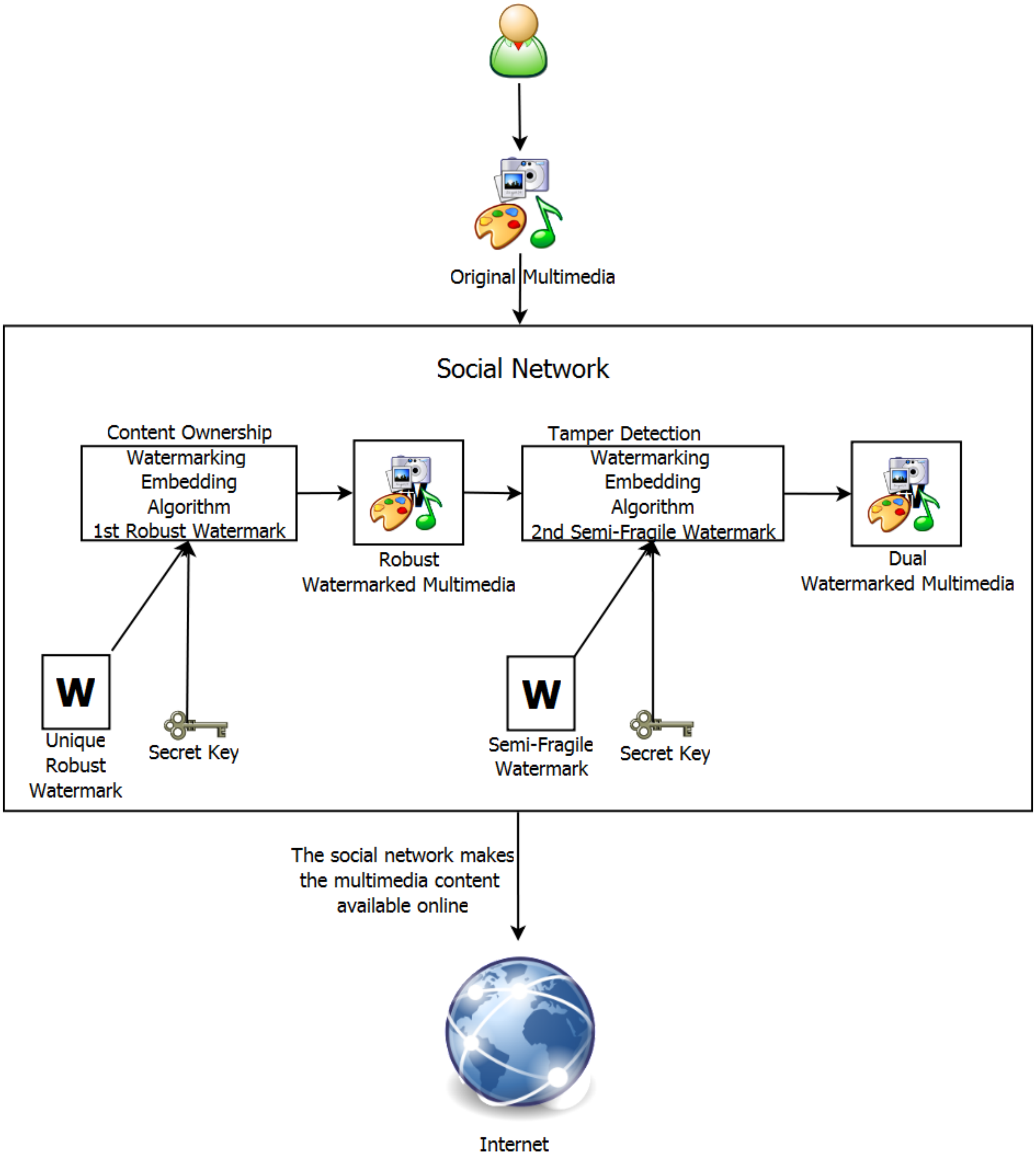

Original Multimedia
Social Network
Content Ownership Watermarking Embedding Algorithm 1st Robust Watermark
Robust Watermarked Multimedia
Tamper Detection Watermarking Embedding Algorithm 2nd Semi-Fragile Watermark
Dual Watermarked Multimedia
W
Unique Robust Watermark
Secret Key
W
Semi-Fragile Watermark
Secret Key
The social network makes the multimedia content available online
Internet

Figure 6.4: Dual watermarking Scheme

Let's assume that user A is the owner of some original multimedia content and he decides to upload it to an OSN. Then the OSN embeds the dual watermark to the uploaded content and makes it available to other users according to user's privacy settings, e.g. public, friends of friends etc. The User B is a user that has the right to access the multimedia content and since a user can access it he can download it and store it to his computer. The link that user B has for the content is not static. The use of dynamic links for contents in OSNs must be used at any time, as the static links are the most obvious way that users can bypass any sort of privacy policy. The case now is "you see it once you own it forever" no matter if one changes his privacy policy. Moreover, a static link is easy to be copied and shared not only inside a OSN, but in the whole Internet as well.

Now let's suppose that user B wishes to re-upload the multimedia, to the same OSN, with or without making any changes to it. The submitted file is checked from the OSN for the existence of the first watermark, the robust, which shows the owner of the content. When the OSN see that the robust watermark is present then checks the privacy settings of the user A, who has the ownership and according to it, the content is allowed or not to proceed to next check. If user B has the privileges to repost the content then the last step is to check the integrity of the content with the semi-fragile watermark. If the content has been tampered, then an alarm is triggered for user A, showing the altered version of his content requiring his consent for resubmission. In order to avoid possible problems, a logical timeframe for this answer is being applied, so that if user A hasn't answered for e.g. a month then this means that he doesn't care for this post, therefore it is automatically published. In case user B has the right to submit the file, user A just receives a notification for the event. This workflow is illustrated in Figure 6.5.

A different approach would be embedding of watermarks on the fly, every time an authorized user is granted access to the content, Figure 6.6. This allows the use of non-blind algorithms, which are more robust, while enhancing the watermark system with fingerprinting capabilities. On the fly watermarks could include the user ID of the user that gains access on the content. Therefore if the content leakage has been made by some user to another SN or the Internet generally, it can be traced and settled more easily. The obvious trade-off of this approach is the computational cost on the server side.

Although OSNs regard themselves as completely separate worlds that do not interact, this is not the case, as there is a direct link between all of them and this is no other but their users. More precisely, the users which have accounts to other OSNs. The idea of "one OSN to rule them all" does seem probable, as more or less each OSN has its own target group, providing different functionality and services to its users. Given that a

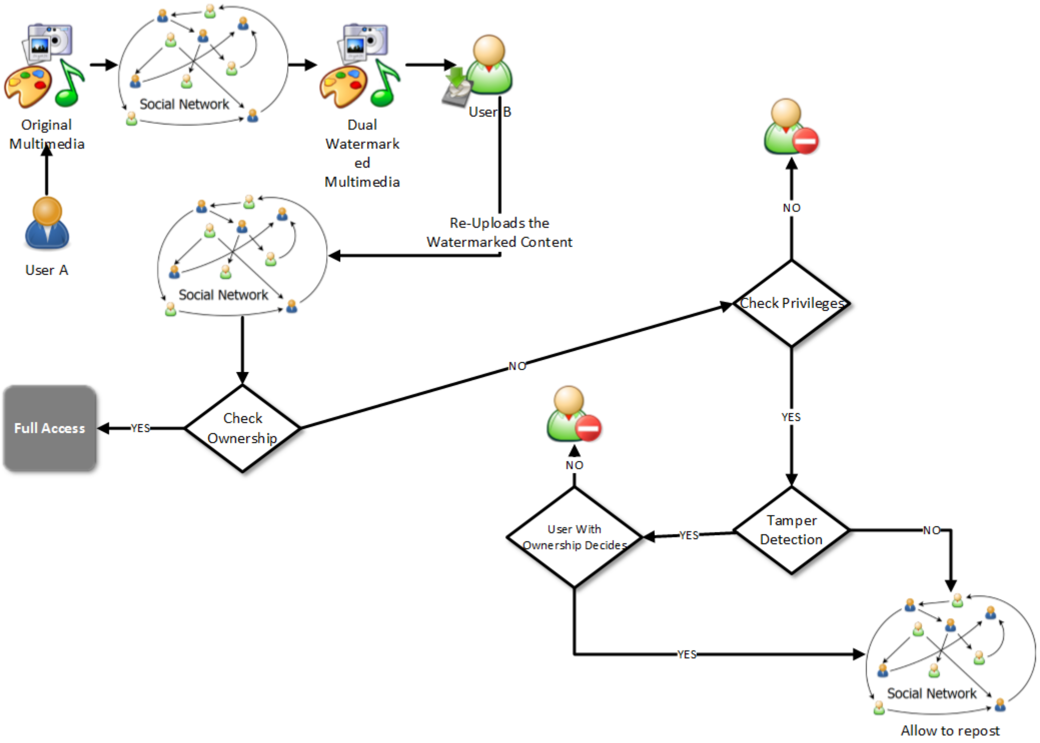


Figure 6.5: The Zigomitros *et al.* scheme.

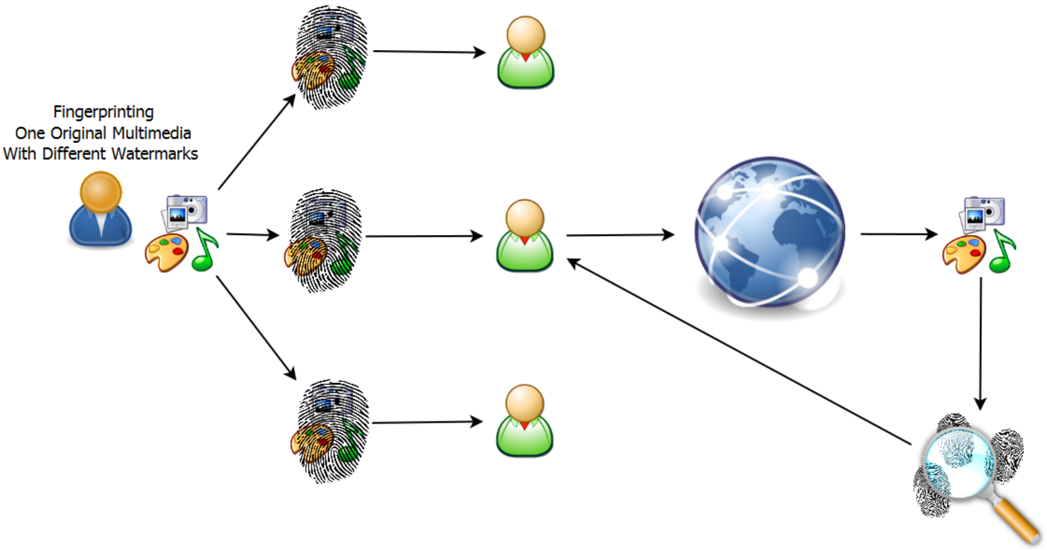


Figure 6.6: Watermarking for Fingerprinting.

unification of OSN's is not probable, the only solution to enforcing privacy measurements across multiple OSNs is their cooperation. The scheme as depicted in Figure 6.4 can easily applied to a single OSN, however extending it across multiple OSNs would result in several problems. The most obvious one is who is applying the watermark. Creating a Trusted Third Party (TTP) which watermarks every medium that is supplied from the OSNs might sound a good idea. However, this would demand the generation of new data centers and additional communication costs. Nevertheless, the major drawback is the fact that OSNs would have to go under the umbrella of a unique authority. The latter could be accepted from small OSNs, however, major OSNs are highly impossible to accept such an approach, given their market position. A solution without a TTP and with minimal interaction between OSNs is feasible. This can be achieved by altering the watermark that each OSN applies to the uploaded media. For the feasibility of the solution we could assume that there are n OSNs that cooperate on enforcing cross-OSN privacy policies on multimedia content, that share a common watermarking key K[4]. Also without loss of generality, it must be assumed that a user uploads an image file, however, the procedure is the same for any multimedia file. We estimate that the least information that should be embedded in each watermark is the following: The userID which allows each OSN to determine the owner of the media. A mediaID field which notifies the OSN where the image was originally hosted. There is also a need for a timestamp field to indicate when the media was watermarked. Finally a publication license ID is also needed. This information might seem unnecessary, as an OSN will have to check for the user's policy. Nevertheless, this may solve other problems that are going to be discussed in the following paragraphs. Moreover, each OSN has its own private and public key pair $(Priv_{OSN_i}, Pub_{OSN_i}), i \in \{1, ..., n\}$ and a symmetric key Sym_{OSN_i}.

If that Alice uploads an image to OSN_1, then OSN_1 creates a vector v as follows:

$v = \big(E_{Sym_{OSN_1}}(UserID||rnd), MediaID, Timestamp,$
$PublicationLicense,$
$E_{Pub_{OSN_i}}(OSN_{m_1}Data), ..., E_{Pub_{OSN_i}}(OSN_{m_k}Data)\big)$
where:$\{m_1, ...m_k\} \subseteq \{1...n\}$ and rnd a random value.

The first field is encrypted with Sym_{OSN_1} so that OSN_1 can recover the UserID quickly. UserIDs are salted with a random value in order to obfuscate the UserID. Leaving the userID just encrypted allows other OSNs to profile users by storing the encrypted form of their IDs. If they are salted, then only the original OSN can find the owner of the media and all other OSNs are blinded not only from the owner, but from any other media of the same user. The next three fields are not encrypted, so that everyone can retrieve the mediaID, the timestamp and the publication license of the

[4] K is used to watermark each image with a dual watermark, a robust and a semi-fragile as in the original Zigomitros *et al.* scheme.

user. Finally, the rest of the fields contain information that is specific for each OSN and can be retrieved only by them. The vector is signed by OSN_1 so the information that is embedded in the watermark w is $w = v, E_{Priv_{OSN_1}}(H(v))$, where H is a secure hash function. Using K, OSN_1 embeds the dual watermark in the photo and publishes it. When a user wants to upload the same photo to OSN_2, then OSN_2 will use K to extract the watermark. From that, OSN_2 will get the vector w and verify that it is correctly received from the digital signature. Based on the publication license and the message that OSN_1 has encrypted for OSN_2, OSN_2 will decide whether or not it will publish the photo and with what privacy settings, notifying OSN_1 about these actions.

6.2.5 Discussion

The proposed scheme enables users to apply their privacy policies on their multimedia content across multiple OSNs. It is important to highlight that the users do not need to be registered to all OSNs to allow this functionality. Moreover, users can be notified of any attempts to violate their privacy. The scheme does not need any trusted third party, therefore, there is no further trust dependency.

The use of timestamp in watermarks is considered essential as they can be used to define fine-grained policies in our scheme. Since each photo is watermarked on upload, users can use this information to define time-based policies. For instance, a user may allow a photo to be public after 2 years, or stop sharing one after 5 years. On top of that, timestamps can be used in case of conflict to determine which user has uploaded the content first and deduce its origin. The latter can be understood only in the case when a new OSN joins and checks its content against its peers[5].

The introduced publication license field is very important, as users may use standard licenses such as Creative Commons[6] or define custom ones, excluding specific users or OSNs from distributing the content. It is clear that personal photos will have custom policies, while others will have more generic ones. To illustrate this concept, we assume that Alice publishes a photo with a “Creative Commons Attribution - Non-Commercial - ShareAlike 3.0 Unported License”. This means that Alice does not allow modifications of her work or commercial use. Bob finds this photo and can publish it in his profile. If Alice decides to withdraw the photo, then this will not have any impact

[5] It should be highlighted that while the proposal is straightforward regarding new content, managing already published content or how a new OSN joins is more complicated and is going to be discussed extensively in future work.

[6] http://creativecommons.org/

on Bob's profile and Bob will not have any problem with publishing the photo even after Alice removes the photo. however, if Bob has downloaded the image, processed it and tries to upload it to OSN_2 where Alice is not registered, OSN_2 will detect the alterations from the dual watermark and since the license does not allow modifications, block Bob from uploading the photo.

Another possible scenario, occurs when Bob might try to upload a photo from Alice's profile in OSN_1, where her characteristics are quite clear into the professional OSN_2 where Alice is not registered. Since Alice's photo is personal, she has watermarked it with a non distribute license. Therefore, if Bob wants to perform an identity theft attack to Alice, OSN_2 will block his actions by reading the embedded watermark.

The scheme allows OSNs to have different policies among them, without publicly disclosing them. Therefore several OSNs, depending on their interests, conflicts and policies, may choose to cooperate under different schemes, without exposing critical information to the rest of the participants. This way, Alice who is registered in OSN_1 can allow only users from OSN_2 to re-upload some of her photos. Given that Alice might have two accounts on different OSNs, she can notify OSN_1 that photos are co-owned by another user from OSN_2, specifying her ID in OSN_2 and vice versa.

The proposed solution reestablishes the roles of OSNs, as not only do they host content, but they become Content Certification Authorities (CCAs). CCAs can certify the origin of a submitted multimedia file, hence detect if it belongs to one of its users or not, to the users of another affiliated OSN and even detect alterations. Evidently, this scheme enables not only privacy aware sharing of media content, but furthermore the unification of user accounts among different OSNs. This unification might seem at first sight scary for most of the OSNs, especially the ones with fewer users. Nevertheless, depending on the differentiation of the services that each of them provides, this unification can only enhance their status. This can be achieved due to the fact that the unification can enable developers and OSNs to deliver more solid, useful and fine-grained solutions to the users. The decentralized nature of the scheme enables the equal treatment of all the participants, which is very crucial for its continuity, creating a web of trust not only among the OSNs, but among their subscribers as well.

Maybe the principal advantage of this scheme is that user's privacy is greatly enhanced, as the user has total control of his media. He can keep track of where his media files are being used, who has access to them and revoke or grant access to them in real time, independently of the OSN that he is registered. The obvious drawback of this solution is what happens with the already shared content and how to tackle cases where different users share the same content and one of them declares ownership. Of course, human intervention cannot be avoided, yet the best approach would be to watermark all this content by current OSNs and mark it as non further distributable, unless all the parties agree on the ownership. Finally, the proposed scheme allows OSNs to automatically respond to changes in the legal system. In the upcoming years, many changes are expected to be made in the privacy laws in national and international level. This may have serious implications for OSNs as they will have to change the way they distribute content according to the new laws. A framework, as the one we propose, allows OSNs to automatically conform, as the changes in one of them will lead to cascading changes to the rest of the OSNs, significantly reducing the cost of law compliance.

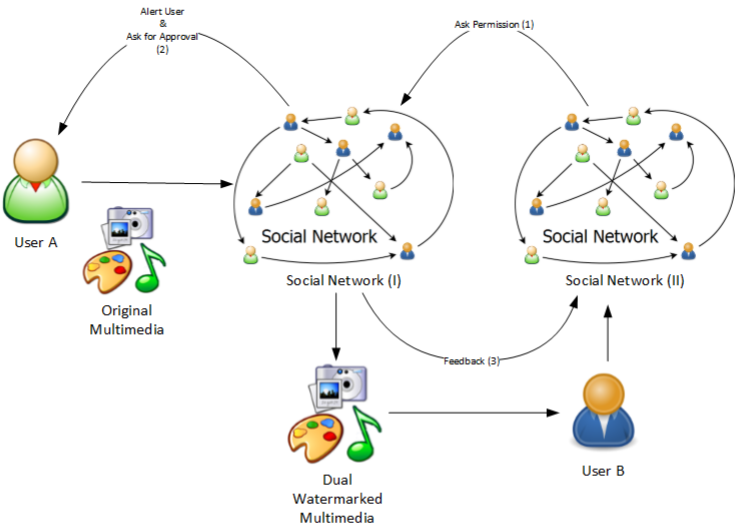


Figure 6.7: Managing media files in two Social Networks.

6.3 Economical impact

Almost every OSNs operate under the so-called "freemium" model, meaning that they offer the service free to the users, in exchange for accessing, mining their data and offering targeted advertisement to their customers. Watermarking their massive amounts of photos is certainly a serious cost, as all this procedure demands many additional processing hours. Therefore a very important question that has to be addressed to is the economical impact of the proposal in the current established business model.

The economical impact of the proposed solution depends directly on the computational effort is needed to apply this scheme. To estimate it, a Java implementation of a typical DCT watermarking scheme. It was used on a desktop computer running on Intel Core i7-3770 CPU clocked at 3.40GHz and without a fully optimized implementation, a photograph with a resolution of 2048x1536 needs on average approximately 1.5 sec to be watermarked per core. Given that this processor has 8 cores, such a computer can watermark around 460800 photos per day. According to [111], each day, 350 million photos are uploaded to Facebook. Therefore, Facebook would need approximately 760 such computers to balance the computational effort for its daily traffic. On the other hand, it takes around 1 sec to extract the watermark. It should be highlighted that these figures could be further reduced with a more optimized implementation or with exploiting GPUs.

The cost of maintaining this additional infrastructure cannot be considered nor negligible nor prohibitive, nevertheless is something manageable. Currently, even if millions of people are using OSNs, sharing huge amounts of information, they are aware that this way is not the most privacy-aware method. If some OSNs decide to build or reformat their structure, offering more privacy to their users via their collaboration, on the one hand, they will increase their maintenance expenses, on the other hand, it is expected that they will attract many additional revenues. Firstly, providing a feature such as cross-OSN privacy policies is expected to attract more users, especially at this point in time where people are becoming more aware of privacy. The latter was sparked by recent relations about background actions from secret government agencies. Many start-up companies are rushing to exploit this new niche market of privacy. Therefore, since the proposed scheme minimizes the leakage of users' information, many new users will be attracted. Moreover, people will be able to share more information, or even more sensitive, as the privacy-aware shift from these OSNs will renew their trust to the service. All the above, make the collaborating OSNs more attractive to advertising companies, as they will host more people, more information to mine and perhaps more valuable for the advertising companies.

A privacy-aware service for subscribers is an additional feature that can attract artists of all genres to publish and share more of their work on OSNs. With this target group, but without watermarks, Pheed[7] is gradually getting more and more users promoting itself as a free social multimedia platform that allows multimedia sharing and streaming. Given that many artists do not share their work due to leakages, the proposed scheme would make such OSNs even more attractive. On other hand, watermarking techniques are considered essential, as they provide the necessary trust to the users that access them. Typical examples are professional, medical and dating OSNs or their extensions. In such cases, it is crucial to guarantee that the users

[7] www.pheed.com

are real and that their profiles contain the right information. Identity theft or even extortion attacks on such cases can harm the public and professional image of the victim severe and immediately. To address such challenges, applications such as Badoo[8] have resulted in visible watermarks, degrading the actual medium. However, the necessary functionality is not provided, as these watermarks could be easily cropped in many cases. Consequently, it is clear that subscriptions of premium accounts or the attraction of more users and better quality of content sharing can mitigate the costs of applying watermarking schemes in OSNs and maintaining the needed infrastructure.

[8] www.badoo.com

This page is intentionally left blank.

Chapter 7

Privacy and Security in Mobile Applications

7.1 Introduction

The technological advantages of ICT radically transformed the way we communicate. Up to recently, our interactions were bounded to our immediate surroundings, but nowadays the distance factor has been eradicated.

Mobile technology reaches a huge market and smartphones are accessible and affordable in almost all parts of the globe. Their continuous improving computing capabilities merged with sensors like accelerometers and GPS gave to software developers the opportunity to deploy applications that exploit spatial information. As a result of these new means of interaction, users are more engaged and developers can make them actively participate and provide added-value content and information.

The thriving mobile dating applications are a good example of these communication and behavioural changes. These applications changed drastically the way people are looking for partners. Due to their very nature, dating applications contain sensitive information and the use of these does not come without risks. In addition to the sexual orientation and preferences of users, modern dating apps are also context-aware and allow the users to find near-by matches. Moreover, the usersś profile may contain more sensitive information such as political views and beliefs.

Users choose to share some personal information with potential partners as it can be used as filters for achieving better matches. Notwithstanding, users might not be comfortable this kind of information to be publicly available. Moreover, the exchanged information between two users is very sensitive and private.Thus, any leaked information from the applications could damage the reputation of individuals.

These privacy and security risks are well-known and defined and someone could expect that the developers are taking the appropriate measures to secure their applications and to avoid any privacy breach, especially since their applications have millions of users.

We conduct a study on 18 mobile dating and chatting apps and we highlight some doubtful software development practices which seem to be commonplace. The vulnerabilities we have detected, in many cases, are quite obvious and so are their solutions. Nevertheless, these vulnerabilities affects millions of users and in a variety of ways. The most worrying observation was that the exploitation of these vulnerabilities requires little, if any, computer skills. In our experiments, the most *sophisticated* scenario implies to intercept network packages *e.g.* by using a proxy, while in the simplest scenario, the attacker just needs to eavesdrop exchanged messages.

7.2 Related work on Security and Privacy in Mobile Applications

The wide adoption of Internet technologies has led to the development of novel services. The latter does not come without a cost, as attackers may now perform attacks remotely and affect millions of users. Many organizations strive to raise awareness against these attacks and for a secure web development. For instance, OWASP[1] compiles the well-known OWASP Top Ten survey, which highlights the most critical web application security flaws. This survey provides a very good insight on what developers should be aware of when developing applications and it illustrates how attackers would try to penetrate into a web service. Other surveys such as [180, 39] provide similar results.

Mobile applications usually are not developed as stand-alone services which run solely on the mobile device, but rely on web content and infrastructure. In this case, the content is retrieved, updated and uploaded through the Internet. The mobile's sensors gather data that are used to provide value-added services and functionalities.

Currently, the dominant operating systems for smartphones are Android and iOS. Android offers a permission-based security model, however, as has been shown in [97, 11] it has vulnerabilities. Beresford et al. [22] developed the *MockDroid*, which is a modified version of Android and allows the user to intervene and revoke the access rights of applications to particular resources at run-time. Certainly, this approach has a very negative effect on the usability of the application because it limits some

[1] https://www.owasp.org

functions. Notwithstanding, it grants users to set their own privacy policies and assist them to decrease their possible information leakage. Having a similar perspective *TISSA* [296] enables the users to fine-tune their privacy policies and the access level of specific applications at run-time. Enck et al. had a different approach to the problem when introduced *TaintDroid* [70]. Instead of constantly intervening, their Android modded version tracks down when and which sensitive data is requested by an application, hence, providing real-time analytics. This analysis can be very helpful in the automated characterization of benign, grayware and malware [71].

In the iOS ecosystem, publishing an application in the official app store passes some stricter filters. Egele et al. [64] made an extensive study over a sample of 1,400 iOS applications (3/5 from the iTunes App Store and 2/5 from the Cydia repository). While most applications were found to be benign, their analysis revealed that more than 50% of them was leaking the unique ID of the mobile device.

Wetherall et al. [270] developed two useful tools, a browser plugin and a mobile application, in order to provide to users a clear overview of the dangerous information flow in their mobile devices. These tools warn the users on unencrypted logins and inform them which sensitive information is collected.

In [31] Burattin et al. show how can an information, e.g. list of friends, that have been set to private by the user to be recovered from an OSN with various methods. Focusing on dating applications, Qin et al. [201] illustrates how an adversary could obtain the geolocation of users in several widely used dating apps. As shown by Qin et al. [201] the main reason for this exposure is the poor randomization and obfuscation techniques used.

The Electronic Frontier Foundation published a review[2] to determine the kind of security is offered by secure messaging products. The review clearly indicated that there were many problems in the architecture of most applications and the user cannot be considered secure as for instance it is impossible for the user to verify the contact's identity, or the service provider has access to the users' traffic.

7.3 Experimental Setup

Our approach to the problem of obtaining sensitive information was from a non-invasive perspective. Taking in consideration the legal implications to applying reverse engineering on an application, and the capacities of an average user or a network administrator, we have installed a proxy to intercepted the messages targeted towards to the

[2] https://www.eff.org/secure-messaging-scorecard

under review applications. In order to ensure that the content of the intercept packages can be analysed by the proxy, we generated a root certificate for it and installed it on the smartphone. By doing this even the encrypted packages could be read by the proxy. This setup resembles the *man-in-the-middle attack*, with the only, but very important, difference that all sides are under our control.

There are two real-world attacks scenarios that can be launched based on the above setup. The first scenario involves a network administrator that is curious to gather as much information as possible about the users in the network he controls. There is no need to request users to install a certificate because many applications are using unencrypted traffic or leave important information in the header so the administrator only needs to watch regular network traffic. In general, anyone sniffing the networkś traffic can execute these attacks.

The second scenario assumes that a malicious user has cyber-stalking intentions. The adversary wants to find probable victims and their whereabouts. To achieve this, he intercepts the packages that are sent and received from the applications and then uses them to extract further fine-grained information. Figure 7.1 illustrates our setup and where we expect our adversary to be logically located.

In order to capture the data from the applications we have used Fiddler 2 [3] as the debugging proxy and we have installed its certificate in a smartphone. For our experiments we used an iPhone 4 with iOS 7.0.4. However, the operating system is irrelevant since the vulnerabilities we have found are mainly due to the exchanged traffic between the applications and web servers that host the services and, in fact, the applications *per se* have not been analysed.

7.4 The findings

In our experiment, we analysed 18 mobile dating and chatting applications and examined whether they send sensitive HTTP traffic, include the current geolocation of users, send the actual distance to other users, use static links for multimedia content, etc. The Figure 7.2 summarize our findings. Follows a brief description of specific details/vulnerabilities of each application we examined.

ChatOn: ChatOn uses HTTPS for all its traffic. However, a lot of information could be leaked through the URL of the API. For instance, an eavesdropper can easily find many details about the user's mobile device, namely, model, operating system version, IMEI,

[3] http://www.telerik.com/fiddler

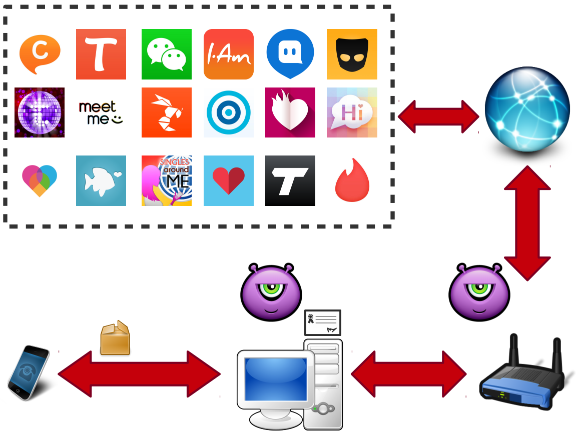


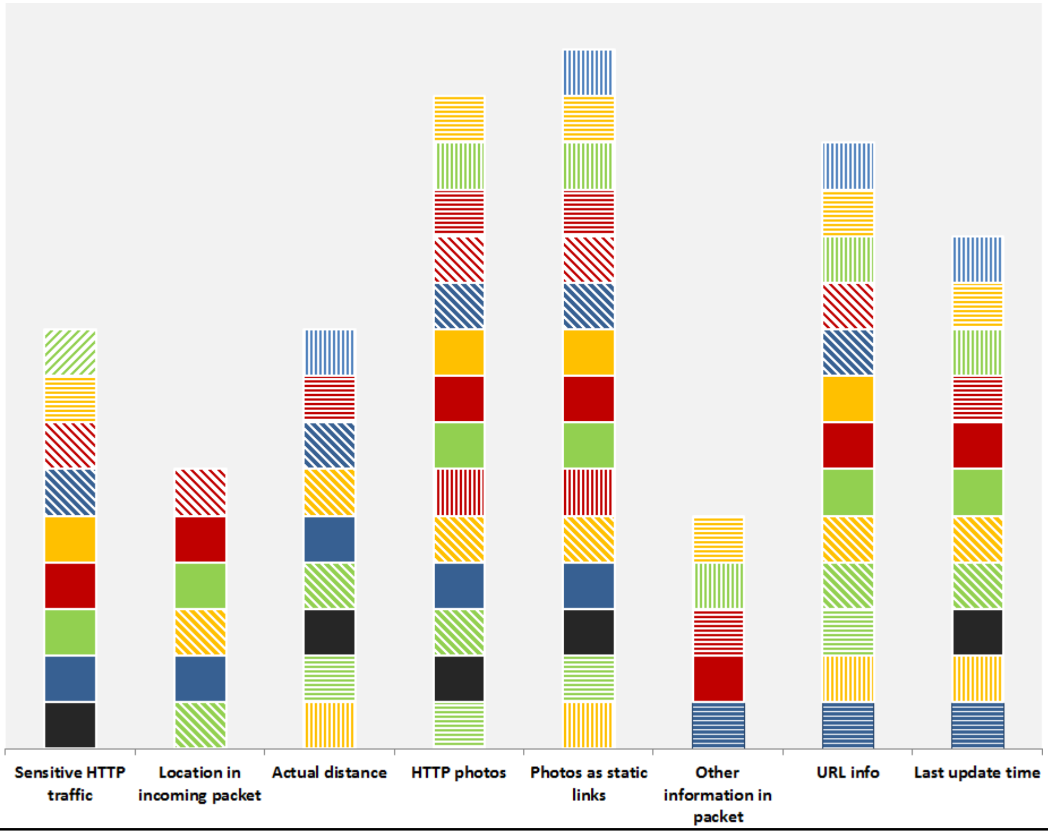
Figure 7.1: Experimental setup

IMSI, telephone number, user ID and application version. The application sends the telephone numbers of all user's contacts to Samsung, and the received packets contain additional information like contacts' birthday. The RESTful API that is used exposes users actions and the profiles that they visit.

Grindr: Grindr uses HTTPS for most of its communications, but multimedia content are sent by using static links over HTTP. The API that is called from the mobile application might allow eavesdroppers to extract the actual user location and his/her application ID from the sniffed URL. Additionally, the URL reveals the user's activity and his/her device OS. Moreover, exchanged packets contain the distance only for users that consented and the application might display the relative user distance. However, the messages contain the actual users' location.

Hornet: Hornet encrypts its traffic using HTTPS, but sends the distance with 10 metres precision. Photographs are static links sent over HTTP. The API calls allow an attacker to deduce user activity, e.g. chatting, browsing profiles, etc simply by capturing the URLs that users request.

I-Am: Only the authentication executed over HTTPS, the rest of the traffic of I-Am goes over HTTP. This allows an attacker to have full access to user private data. Photographs are sent over HTTP and as static links. The application exposes the spatial data of users since the actual location is sent in the URL without any encryption or obfuscation, and the exact distance to other users is sent in the packet along with their birthday.



Application	Version	Installations	Code	Application	Version	Installations	Code
ChatOn	3.0.2	100m-500m	▤	S.A.M.	3.3.1	500K-1m	🟥
Grindr	2.0.24	5m-10m	▥	SKOUT	4.4.2	10m-50m	▧
Hornet	2.0.14	1m-5m	▤	Tagged	7.3.0	10m-50m	🟨
I-Am	3.2	500K-1m	⬛	Tango	5.8	100m-500m	▧
LOVOO	2.5.6	10m-50m	▧	Tinder	4.0.9	10m-50m	▤
MeetMe	9.2.0	10m-50m	🟦	Tingle	1.12	-	▥
MoMo	5.4	1m-5m	▧	Waplog	3.0.3	5m-10m	▤
POF	2.40	10m-50m	▥	WeChat	6.0.1	100m-500m	▨
SayHi	3.6	10m-50m	🟩	Zoosk	8.6.21	10m-50m	▥

Figure 7.2: Discovered vulnerabilities per application. Since the App Store is not reporting the downloads, the numbers are from Google Play.

LOVOO: The photographs in LOVOO are sent over HTTP, everything else are sent over HTTPS. Something worth noting is that the links are dynamic and expire. The distance between users is sent with a rounding of 100 metres, along with their relative (x, y) coordinates. Thus, an attacker can recover the actual locations. Moreover, the API calls expose in the URLs the userś preferences, location and overall activity

MeetMe: A mix of HTTP/HTTPS traffic are used in MeetMe application. In URL exposed the location and the preferences of users. If other users are nearby then he actual distance is included in the packet, otherwise their distance are reported in kilometres. Photographs transmitted over HTTP and the user actions can be seen in the URLs.

MoMo: MoMo uses HTTPS to exchange messages with the server, but fails to hide users' location. The application packets contain fine-grained distance data from other users. Moreover, the URLs include the visited profiles as well as the current user ID. Photographs are accessed through static links and over HTTP.

Plenty of Fish: This application messages are encrypted although they are transmitted over HTTP. This most likely means that there is a local key stored in the device to decrypt the contains. They did not take any precaution for the multimedia content which are send over HTTP as static links.

SayHi: The user authentication on the application use the Facebook API. After the authentication, everything is transmitted as clear text. The location of users and their activity are exposed in the requested URLs. Even the userś conversations can be intercepted. Photographs are sent over HTTP using static links.

Singles around me (S.A.M.): The exact location of other users are included in the transmitted packets. Multimedia content are sent over HTTP with some of them have dynamic links and others static. The received packet includes an important additional field: users' emails. Usually, a user needs to ask for permission to access another user email address and not to be able to extract it from the packets. Moreover, the URLs contains the IDs that a user visiting inside the application, thus reveals the activity and his/her preferences.

SKOUT: Authentication of users done with HTTPS but the rest of the traffic is sent over HTTP. The exact distance among users are in the packet and only obfuscated in the frontend of the application. Since traffic is sent over HTTP, chat messages are unencrypted. and it is easy to check the activity of a user, e.g. there is a specific call of function whenever a user is typing a message. Anyone who can intercept the traffic can find not only the sexual orientation of a specific user but his/her preferences as the profiles of the user that he/she opens are visible to the adversary.

Tagged: Everything is sent over HTTP, therefore all messages can be intercepted and the API exposes user's activity and preferences. Photographs are sent over HTTP as static links.

Tango: Tango transmits over HTTP and all messages can be intercepted. The API exposes user's activity as well as his/her phone number and preferences. Photographs are sent over HTTP as static links and the packets include the exact location of other users.

Tinder: Tinder uses HTTPS to prevent eavesdropping but the messages contain the Facebook ID of the users. Therefore, an adversary can find out even more information. Packets do not contain the current location but the distance to other users, which can be further exploited to recover their actual locations. Photographs are sent over HTTP as static links. In Tinder despite the use of static links, we noticed an additional privacy leakage when using its registration method which is via Facebook. The application was requesting the "communication_rank" of each user from Facebook and it was sending it to the application. As the name of the parameter implies this variable indicated how often two users communicate over Facebook.

Tingle: Tingle does not include location data of other users in the transmitted packets. However, messages contain other important information. Like Singles Around Me, Tingle exposes other users' emails and, additionally, it has a device tag indicating, for example, that the user has switched the device. Moreover, it contains the actual location of the user in the URL, allowing an eavesdropper to identify him/her. The URL used by the API contains users' queries, hence, exposing their preferences. Finally, photographs are sent over HTTP as static links.

Waplog: Waplog transmits over HTTP. While it does not send the actual location, it transmits emails of other users. Photographs are sent over HTTP as static links. In addition, the API exposes information about the user's device in the URL, the session key and the user's hashed password.

WeChat: A different approach from the WeChat application which uses HTTP for all its traffic and but it sends all data in an encrypted file. During our experiments, we didn't notice any handshake between the application and the API to generate a cryptographic key in order to decrypt the encrypted file. Consequently, we may safely deduce that the application is installed with a static hard-coded key, which is the same for each user and with reverse engineering of the application can be recovered. Therefore, one could use that key to fully manipulate the data of all users.

Zoosk: Zoosk transmits over HTTPS. The requested URLs reveals the phone model and its OS as well as the user activity. Photographs are sent as static links over HTTPS.

7.5 Discussion

The analysis of the above mobile dating and chatting application revealed several trends that we are going to discuss next.

Users' locations: The exact location of the users is handed over by many of the studied applications in order to match nearby users. The Table 7.1 illustrates a typical example where the application "Singles Around Me" sends a JSON file which includes the GPS location of a user. This is a problematic approach and susceptible to many attacks. To avoid to transmit the exact location other applications use the distance between users. It seems to be a better solution but still vulnerable to trilateration as shown by Qin et al. in [201]. The developers should examine the use of *Private Proximity Testing* protocols such as [169] which reduce the users' exposure while the application keeps the desired functionality. In these protocols, only a single bit of information is disclosed, *i.e.* whether two users are nearby or not.

Unencrypted transmission channels: As shown in [74] many mobile applications use SSL/TLS code, which is potentially vulnerable to man-in-the-middle attacks. Notwithstanding, these applications were, at least, trying to protect the users from exposure. In our analysis, major applications use HTTP to transmit sensitive data. The use of

```
{
 "username":"s---------eam",
 "email":"d----------96@yahoo.com",
 "gender":2,
 "interestedIn":1,
 "country":"United Kingdom",
 "region":"London, City of",
 "city":"",
 "gps":[38.------------2,23.8-------------5],
 "age":39,
 "photo":"http://www.singlesaroundme.com/images/cache/1/photoface_11-----_--5_--5.jpg",
 "photos":[],
 "birthYear":1974,
 "birthMonth":--,
 "birthDay":-,
 "lastOnline":"2014-10-06 03:28:07PM",
 "profileAnswers":{"1":"5' 7" - 170cm",
 "3":"prefer not to say",
 "21":"Married","30":"straight","25":"brown","31":"blonde","26":"white","28":"none",
 "29":"Sagittarius","38":["dating","serious relationship","friendship"],
 "37":"Greek","36":["English"],"32":"socially / occasionally",
 "34":"socially/occasionally","35":"quite fit","40":"Christian - Other",
 "41":"University graduate","42":"yes living with me","43":"yes"},
 "privacySettings":{"gps":0,"profile":0}
}
```

Table 7.1: Example of a JSON packet from “Singles Around Me”.

HTTPS on mobile devices might imply an overhead in terms of computation and bandwidth. However, these applications manage very sensitive information and it is difficult to support the use of HTTP. Given that the overhead of enforcing HTTPS is negligible and that converting services into running in secure mode is most of the times a matter of adjusting the configuration implies that developers and architects are not aware of the actual risks they expose users to. Finally, we highlight the lack of full adoption of HTTPS. Despite recent revelations about citizen surveillance or recent cyber attacks, services that are used by millions still fail to use HTTPS, they offer it partially exposing their users to disclosure of very private information and profiling.

Multimedia: The applications we examined performed poorly on the handling of multimedia content. The privacy and security risks of multimedia content on Online Social Networks were analysed by Patsakis et al. in [189]. Multimedia content on mobile dating applications could be considered even more sensitive and implies further dangers. The lack of use of HTTPS allows an adversary to intercept and even change the content of the incoming messages. Moreover, the use of static links is a serious vulnerability because an eavesdropper can access the photographs which a user has visited. By doing so the adversary can identify the preferences and sexual orientation of users. Such information, as sexual orientation and preferences, may lead to social discrimination or even legal actions in a closed society.

Hidden information and URL parameters: In our analysis, we noticed that a common practice among several applications was to send data packets which include hidden information about other users. An example of this practice in application "Singles Around Me" illustrated in Table 7.1 where the email of users are revealed even if they have not consented. Tingle packets contain a device identification field which informs an attacker when a user changes or switch devices. Another observation of our research was that sensitive parameters are passed in the URLs of the API. Simply by eavesdropping the communications, an adversary can monitor the users' activities (*e.g.* browsing of profiles, chatting, etc), the telephone number or even the location of the user. A small set of examples that leaks information through URL are illustrated in Table 7.2. In all these examples the traffic is encrypted, but still, someone can see that ChatOn broadcasts the IMEI, IMSI and user's phone number in the URL along with some details about the phone, Grindr broadcasts the user's location his ID and some data about the device. In the same way, MoMo leaks which profiles a user is visiting and SKOUT shows what the user is doing, in this case typing.

Application	URL
ChatOn	https://gld1.samsungchaton.com/prov4?imei=-----&imsi=-----&model=iPhone4&clientversion=3.0.2&platform=iPhone%20OS&osversion=7.0.4
	https://prov4?imei=------&countrycallingcode=30&phonenumber=-----&imsi=----&model=iPhone4&clientversion=3.0.2&platform=iPhone%20OS&osversion=7.0.4
Grindr	https://primus.grindr.com/2.0/broadcastMessages?applicationVersion=2.0.24&hasXtra=0&lat=53.-----&lon=6.2----&platformName=iOS&platformVersion=7.0.4&profileId=36850131
MoMo	https://api.immomo.com/api/profile/1121----?fr=98----
SKOUT	http://i22.skout.com/services/ServerService/GCUserTyping

Table 7.2: User exposure from the URL. For obvious reasons, the sensitive information has been suppressed.

Part IV

Closure

This page is intentionally left blank.

Chapter 8

Open Questions and Future Directions

During the past few years the requests for data sources have been significantly increased. To this end organisations more and more are pushed towards sharing part of their underlying datasets, to be used by other companies and researchers. In order to respond to these requests without exposing users' privacy to further risks, the need for privacy-preserving methods is becoming more and more imminent. To monetise this need, companies such as Apple tout the use of differential privacy in their products [193], or companies such as Aircloak[1] whose core business is the provision of anonymised datasets, are gradually emerging, paving the way for the greater adoption of PPDP. Regardless of the achievements in the field of PPDP, there are several open issues that need to be addressed in the years to come.

One of the biggest challenges that we face today is the management of Big Data, as its three Vs, namely Volume, Variety and Velocity, imply many constraints for PPDP. Starting with Volume, the most obvious characteristic of Big Data, it is apparent that most PPDP algorithms would face issues processing all this information, as these algorithms are quite demanding in terms of computations as their computational complexity, discussed in the previous section, is far from linear. To counter this obstacle, many researchers result to micro-aggregation, while there is a big shift in trying to unravel the potentials of Differential privacy [143, 75]. Certainly, the Variety of data imply another constraint for anonymisation algorithms, since methods as Generalisation depend on such variations. However, the biggest issue is the third V of Big Data, the Velocity at which this information is stored. Undoubtedly, if the velocity is high and data need to be anonymised on the fly, then the only viable solutions can be currently expected from Differential Privacy, nonetheless, the question is whether sequential publications need to be made. In this regard, it has not been studied in depth how

[1] https://www.aircloak.com/

secure the anonymised data will remain when studying analysing publications of data anonymised with Differential Privacy, something that might be subject to the underlying implementation.

Differential Privacy has a lot of potentials in PPDP. Nonetheless, it is far from being considered a panacea or immune to attacks. For instance, Clifton and Tassa [43] have already criticised widespread belief that differential privacy is resistant to attacks, something that was practically shown e.g. by Cormode [44]. Moreover, the noise addition in many cases is subject to the queries that we expect to be performed on the dataset, or whether we have sequential data publications. Some recent works have been proposed towards the generation of incremental ϵ-differentially private releases of data, in order to deal with today's demand for up-to-date information [205, 292]. In this novel research field, the main drawbacks of ϵ-differential privacy, such as the amount of noise required to be achieved are exacerbated and, thus, the creation of more efficient protocols and algorithms is mandatory. In this scenario, the incremental privacy breach, that arises when different anonymised versions of the same dataset are released, has to be taken into account. While several solutions have been proposed in the literature [33, 32, 194, 228], there is still a big gap to fill. Additionally, machine learning algorithms can be used to extract further knowledge from the anonymised dataset [115]. All the above indicate that Differential Privacy might be very useful, however, the privacy guarantees, as well implementations and use case scenarios have to be studied further.

A very active field in PPDP is also the study of high-dimensional data [90]. The curse of dimensionality has a significant impact to the k-anonymity model [7] because most of the data have to be suppressed leading to an unacceptable increase of information loss, rendering the anonymised dataset useless. Currently, high-dimensional data are not found only in healthcare, where traditionally PPDP is applied, but due to the wide collection of data from mobile sensors, this need becomes more relevant to other fields of research and industry. While there are several approaches on how to treat such datasets [162, 84, 267, 200] while more generic approaches try to exploit either the fact that in real datasets the actual background information cannot span to many QIs or the inherent correlation of many QIs [288, 289]. In general, current work in the literature is promising and has paved the way for the future extensions, nonetheless, it is not mature enough for applications and wide adoption, as the data utility of the anonymised datasets is significantly decreased.

With the gradual need to share data and the continuous generation of data from individuals more and more users are becoming publishers and the centralised approach to data publication is shifting to more decentralised approaches. As already discussed, PPDP has already caught up with this trend and specialised methods have been crafted using cryptographic primitives. Nonetheless, these primitives imply a significant cost, especially for large-scale distributed networks [91]. Therefore, secure multiparty protocols need to be improved in terms of computational and communication costs and scalability, to be efficient enough to deal with the PPDP scenarios or more focused protocols on these instances have to develop.

Multivariate datasets have several well-known problems. We consider the proper detection of outlier users (i.e. *outliers*) as a significant issue. In this way, several approaches to deal with outliers in multivariate datasets have been proposed [203, 50, 257, 220, 262]. The presence of outliers in datasets may lead to the disclosure of information about the data distribution and also affect the quality of the obfuscation. Therefore, PPDP algorithms may return biased or inconsistent results due to their presence. For instance, clustering algorithms may incorrectly select users to form a group or a cluster may be wrongly divided into small pieces.

The models discussed so far, however, fail to consider some mining patterns as sensitive information, depending on the context of data. For instance, in the medical field, a relationship between hospitalisation costs and a concrete ZIP area may disclose information that could be used by insurance companies, without disclosing the information of individuals [93]. Therefore, such knowledge patterns must be identified before applying further operations on data or sharing them.

The type and context of data that are going to be protected are very relevant, since it is not the same to protect medical records than, for instance, census. Hence, typical external data sources which could be used in order to attempt re-identification could be used as an advantage to reinforce the data obfuscation procedure. Moreover, for each domain of application, we must identify the specific weaknesses and apply obfuscation so as to efficiently protect data. In this direction, Fu et al. [83] provided a framework to identify which data must be protected and which are most valuable for the utility of the dataset to balance data utility and privacy guarantees.

As future work on the QI Inference attack it should be tested on real datasets and quantify the significance of the attack. As we have shown with the constructed dataset this attack is plausible and may violate privacy guarantees. There is also room for research on how to solve the problem, which is not trivial.

Social Networks can be represented as graphs quite intuitively. The individuals or any other social entity that can be examined, such as a business, are represented as nodes of the graph. The edges of the graph, connect these entities and represent the relationships between them, for instance, people which are “friends”. As shown in [14, 102] removing the identities of the nodes before publishing the graph is analogous to removing the explicit identifiers from tabular data, it cannot be considered enough to anonymise a graph and does not guarantee privacy. An adversary may link a node, from the “anonymised” graph to an individual, by exploiting the structural information of his neighbourhood. The privacy-preserving data publication of OSNs of graphs with the accompanying tabular data while keeping high utility of the anonymised data is quite challenging.

Social networks keep evolving day by day. As new features are added, new attacks on the privacy of the users discovered. Location-aware applications, marketplaces for trading goods, social network specific cryptocurrencies are some recent examples of the ever-changing Social Network landscape. At the same time, even more countries or unions of countries have started to update and publish stricter laws and frameworks regarding the security and privacy of the transmission, sharing and storage of personal data. The E.U. General Data Privacy Regulation (GDPR), the California Consumer Privacy Act (CCPA) and the Brazil General Data Protection Law (LGPD) are some of the latest examples. Furthermore, the upcoming market of the Interent of Things will bring to consumers most of the social features that already use via entirely newly introduced ubiquitous devices enhanced with innovative features and capabilities. The research community must stay vigilant as the privacy threats on early stages of new features are usually severe. As businesses are focused on profits they do not give the proper attention to privacy issues and usually these issues are considered as an obstacle.

Multimedia carries a lot of implicit information and users are not aware of the information retrieval possibilities of an adversary which does not have to be a human, but a machine. Using face recognition, one could potentially scan all the public photographs of an OSN or available on the Internet and find a person of interest within the images even if he/she is not tagged and his/her presence was purely by chance. In the case that someone copies or downloads another’s user image, there is no functionality that can protect them by deleting or destroying this content.

Another issue OSNs fail to implement is the shared ownership over multimedia. It almost impossible to manage content with more than one owner and apply accepted privacy policies by all parties involved. While the problem is real and the content that users share online is significantly increased each year, a mechanism that will be responsible for the shared ownership of the shared files is more than a mandatory option by OSNs and other online content sharing platforms. In fact, the amount of multimedia data that have to be processed and the often lack of origin provides a selter for a lot of deviant behaviour on Social Networks [100, 79, 269, 152, 24]

The research community faces another problem, which is hard to overcome, the identification and handling of malicious accounts. These accounts are used for spreading false news, sybil attacks, phishing, malware distribution and more [82, 215, 78, 259]. Moreover, shifting from centralized to decentralized social networks we must examine the privacy-related issues in the new more complicated environment and examine new issues that emerge.

One of the major challenges is to increase awareness of the users of privacy implications from the use of OSNs which will put pressure on the OSNs to become proactive when dealing with privacy issues. It is worth analysing if there is a need for a user interface redesign of privacy settings on major OSNs, personalized privacy settings recommendations through machine learning based on the clustering of the content and overall behaviour of the users. The education and training of users will also raise the awareness and can be achieved through serious games.

This page is intentionally left blank.

Chapter 9

Conclusions

The implementation of GDPR after many years of negotiations is shaping a new landscape for industries regarding how they collect, process and exchange sensitive user data. This thesis started well before the introduction of this regulations and studied problems and proposed solutions to improve the provided security and privacy.

Prior to a release of a dataset, which contains sensitive attributes to be analysed by the data recipients, it should pass through an anonymisation process. There are many well-known methods to anonymise and each method offers different types of privacy guarantees as analysed in Chapter 2.

Current state of the art failed to notice that the $\mathcal{QI}s$ might be partially dependent on the SA values. In the conducted research a different path than the most of the attacks examined in the current literature was followed. Instead of taking advantage of the $\mathcal{QI}s$, the SA values were exploited to deduce values of generalised $\mathcal{QI}s$. As shown, an adversary is able to break the anonymisation guarantees and infer information of record holders that could lead up to complete re-identification of individuals. The research focuses on medical records because in this sector this phenomenon is quite often. This thesis introduced a new concept, the inference of $\mathcal{QI}s$ attack, and demonstrates the existence of the problem in specific datasets. However, no experiments made on real datasets to quantify its extent.

Additionally, this thesis examined the problem of k^m-anonymising continuous data without using any pre-defined data generalisation hierarchy and a global-recoding heuristic algorithm the `ACD` was introduced. The proposed algorithm greedily selects the optimal generalisation ranges at each step, ensuring all itemsets of size up to m, appear at least k times in the dataset, thus satisfying the k^m-anonymity guarantee. The experiments were conducted with real world datasets and compared to the k^m-anonymisation algorithm `AA` [246] which uses pre-defined generalisation hierarchies.

The results of this approach show that the ACD algorithm, with a small overhead on the computational cost comparing to AA, preserves the utility of the dataset better.

In the third part of this thesis we study the risks to which a user is exposed by his shared multimedia content in terms of security and privacy. Many of these security and privacy risks are indirect or often disregarded by the majority of users. It is shown that even if many actions had been made by OSNs to provide security and privacy to their users, justifiably, it cannot be claimed that the current level is adequate. On the other side, users must understand that they cannot arbitrarily share content with other users and services. This content can be used in many ways, many of which can be proven to be malicious. The problem might not arise from one particular post, but from the fusion of others, or from the background information regarding that post. User awareness through proper notifications might help in this direction, but clearly more media coverage and education can greatly help in this aspect.

We argue that the development of new security and privacy policies for multimedia content is essential. Towards this end, this work introduces a scheme that allows users to enforce their privacy policies not only on multimedia shared in the OSN that they belong to, but among others to which they are not registered. This could be achieved by the use of watermarks on the multimedia with either public encryption algorithms or public watermarking techniques. The major contribution of this work is the unification of privacy policies across multiple OSNs in a distributed way without the use of trusted third parties. Of course, there are other researchers pointing towards the use of watermarking in OSNs [42], yet to the best of our knowledge no experiments on the existence of watermarks in OSNs content have been published, nor has a formal protocol or policy has been implemented by OSNs.

The proposed solution can be implemented without having to redesign current OSNs from scratch, therefore it can be easily adopted, in terms of deployment. One may argue that the proposed methodology hints towards DRM practices, however, the watermarks are only used to protect the users' content and users could opt in or out of this service for all or some of their multimedia content. Additionally, the cost of adopting the proposed solution was quantified in terms of computational effort and solutions proposed to counter the economic cost. As discussed, the main issue towards adopting this solution is the already uploaded content. It is a fact that the implementation of this solution will give ownership to users for all their uploaded content even if they do not have the right to own it. If an image, for example, belongs to user A, yet user B also uploaded it, it would seem that it belongs to user B, so user A should report it in order to settle the dispute. The holder of the original multimedia content can upload it again at any time on the OSN even if a previous watermarked version of this content has been uploaded by another user before the implementation. Definitely, such a scenario is realistic, yet it is apparent that the balance of what can be automated from the proposed solution and what is left on the human factor is drastically decreased, leaving far fewer problems to be manually solved. Nevertheless, the complexity of this issue should be examined thoroughly in future work. Finally, it has to be noted that the longer that such solutions are not applied, the more the cost is increased, as users upload new and more content every day.

When the time came to have hands-on experience with mobile apps, dating apps was chosen as being privacy-sensitive and popular. The number of users of online dating services keeps rising every day. This most certainly will draw the attention of cyber criminals who attempt to acquire sensitive personal data that may be used for user profiling, defamation, blackmailing and even identity theft.

Web-based dating services are well-established and usually the good programming practices are applied. However, the conducted research on dating services on mobile platforms shows that the status there is quite different. One should expect due to the sensitive nature of the data which are collected such as location and sexual preferences, that strong security measures as those implemented in their web-based counterparts would also be present in mobile dating applications. However, as shown in Chapter 7, significant vulnerabilities are waiting to be exploited even by inexperienced adversaries. A simple sniffing attack is enough, for most of the analysed applications, to reveal sensitive information *e.g.* sexual preferences, interaction between users, e-mails etc.

The contribution of this research on the mobile apps is twofold. Firstly, by disclosing these vulnerabilities, the aim is to motivate people about the importance of protecting its privacy and raise awareness to companies offering vulnerable services. Secondly, the solution of these vulnerabilities are usually simple and require a little effort to fix them. Thus, by disclosing these findings the developers are informed to avoid common pitfalls and to aim to secure programming practices in developing mobile dating applications.

This page is intentionally left blank.

Bibliography

[1] DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[2] Uci repository. http://archive.ics.uci.edu/ml/datasets.html.

[3] Uci repository us census data 1990 data set. http://archive.ics.uci.edu/ml/datasets/US+Census+Data+%281990%29.

[4] *2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014*. IEEE Computer Society, 2014.

[5] John M Abowd and Julia Lane. New approaches to confidentiality protection: Synthetic data, remote access and research data centers. In *Privacy in statistical databases*, pages 282–289. Springer, 2004.

[6] Saeed Abu-Nimeh, Thomas M Chen, and Omar Alzubi. Malicious and spam posts in online social networks. *Computer*, 44(9):23–28, 2011.

[7] Charu C. Aggarwal. On k-anonymity and the curse of dimensionality. In *VLDB '05: Proceedings of the 31st international conference on Very large data bases*, pages 901–909. VLDB, 2005.

[8] Charu C Aggarwal and S Yu Philip. A general survey of privacy-preserving data mining models and algorithms. In *Privacy-preserving data mining*, pages 11–52. Springer, 2008.

[9] Dakshi Agrawal and Charu C Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 247–255. ACM, 2001.

[10] Giuseppe Ateniese, Kevin Fu, Matthew Green, and Susan Hohenberger. Improved proxy re-encryption schemes with applications to secure distributed storage. *ACM Transactions on Information and System Security (TISSEC)*, 9(1):1–30, 2006.

[11] Kathy Wain Yee Au, Yi Fan Zhou, Zhen Huang, and David Lie. Pscout: Analyzing the android permission specification. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 217–228. ACM, 2012.

[12] Vanessa Ayala-Rivera, A. Omar Portillo-Dominguez, Liam Murphy, and Christina Thorpe. *COCOA: A Synthetic Data Generator for Testing Anonymization Techniques*, pages 163–177. Springer International Publishing, Cham, 2016.

[13] Julian Backes, Michael Backes, Markus Dürmuth, Sebastian Gerling, and Stefan Lorenz. X-pire!-a digital expiration date for images in social networks. CoRR/1112.2649, 2011.

[14] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW*, 2007.

[15] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. Persona: an online social network with user-defined privacy. In *ACM SIGCOMM Computer Communication Review*, volume 39, pages 135–146. ACM, 2009.

[16] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 273–282. ACM, 2007.

[17] Michael Barbaro, Tom Zeller, and Saul Hansell. A face is exposed for aol searcher no. 4417749. *New York Times*, 2006.

[18] Manuel Barbosa, Alexandre Pinto, and Bruno Gomes. Generically extending anonymization algorithms to deal with successive queries. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1362–1371. ACM, 2012.

[19] Roberto J. Bayardo and Rakesh Agrawal. Data Privacy through Optimal k-Anonymization. In *ICDE*, pages 217–228, 2005.

[20] Filipe Beato, Markulf Kohlweiss, and Karel Wouters. Scramble! your social network data. In Simone Fischer-Hübner and Nicholas Hopper, editors, *Privacy Enhancing Technologies*, volume 6794 of *Lecture Notes in Computer Science*, pages 211–225. Springer Berlin Heidelberg, 2011.

[21] Stefan Bender, Ruth Brand, and Johann Bacher. Re-identifying register data by survey data: an empirical study. *Statistical Journal of the United Nations Economic Commission for Europe*, 18(4):373–381, 2001.

[22] Alastair R Beresford, Andrew Rice, Nicholas Skehin, and Ripduman Sohan. Mockdroid: trading privacy for application functionality on smartphones. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, pages 49–54. ACM, 2011.

[23] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 551–560. ACM, 2009.

[24] Christina Boididou, Stuart E Middleton, Zhiwei Jin, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, and Yiannis Kompatsiaris. Verifying information with multimedia content on twitter. *Multimedia Tools and Applications*, 77(12):15545–15571, 2018.

[25] Joseph Bonneau and Sören Preibusch. The privacy jungle:on the market for data protection in social networks. In Tyler Moore, David Pym, and Christos Ioannidis, editors, *Economics of Information Security and Privacy*, pages 121–167, Boston, MA, 2010. Springer US.

[26] Ruth Brand. Microdata protection through noise addition. In *Inference control in statistical databases*, pages 97–116. Springer, 2002.

[27] Justin Brickell and Vitaly Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 70–78. ACM, 2008.

[28] Garrett Brown, Travis Howe, Micheal Ihbe, Atul Prakash, and Kevin Borders. Social networks and context-aware spam. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, pages 403–412. ACM, 2008.

[29] John S Brownstein, Christopher A Cassa, and Kenneth D Mandl. No place to hide-reverse identification of patients from published maps. *New England Journal of Medicine*, 355(16):1741–1742, 2006.

[30] Yingyi Bu, Ada Wai Chee Fu, Raymond Chi Wing Wong, Lei Chen, and Jiuyong Li. Privacy preserving serial data publishing by role composition. *Proceedings of the VLDB Endowment*, 1(1):845–856, 2008.

[31] Andrea Burattin, Giuseppe Cascavilla, and Mauro Conti. Socialspy: Browsing (supposedly) hidden information in online social networks. *CoRR*, abs/1406.3216, 2014.

[32] Ji-Won Byun, Tiancheng Li, Elisa Bertino, Ninghui Li, and Yonglak Sohn. Privacy-preserving incremental data dissemination. *Journal of Computer Security*, 17(1):43–68, 2009.

[33] Ji-Won Byun, Yonglak Sohn, Elisa Bertino, and Ninghui Li. Secure anonymization for incremental datasets. In *Secure Data Management*, pages 48–63. Springer, 2006.

[34] Isaac Cano, Guillermo Navarro-Arribas, and Vicenç Torra. A new framework to automate constrained microaggregation. In *Proceedings of the ACM first international workshop on Privacy and anonymity for very large databases*, pages 1–8. ACM, 2009.

[35] Jianneng Cao and Panagiotis Karras. Publishing microdata with a robust privacy guarantee. *PVLDB*, 5(11):1388–1399, 2012.

[36] Jianneng Cao, Panagiotis Karras, Panos Kalnis, and Kian-Lee Tan. Sabre: a sensitive attribute bucketization and redistribution framework for t-closeness. *The VLDB Journal*, 20(1):59–81, 2011.

[37] Jianneng Cao, Panagiotis Karras, Chedy Raïssi, and Kian-Lee Tan. ρ-uncertainty: inference-proof transaction anonymization. *Proceedings of the VLDB Endowment*, 3(1-2):1033–1044, 2010.

[38] Aniello Castiglione, Bonaventura D'Alessio, and Alfredo De Santis. Steganography and secure communication on online social networks and online photo sharing. In *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on*, pages 363–368. IEEE, 2011.

[39] Cenzic. Application vulnerability trends report. Technical report, 2014.

[40] Bee-Chung Chen, Daniel Kifer, Kristen LeFevre, and Ashwin Machanavajjhala. Privacy-preserving data publishing. *Foundations and Trends in Databases*, 2(1-2):1–167, 2009.

[41] Sean Chester and Gautam Srivastava. Social network privacy for attribute disclosure attacks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 445–449. IEEE, 2011.

[42] Pawat Chomphoosang, Ping Zhang, Arjan Durresi, and Leonard Barolli. Survey of trust based communications in social networks. In *Network-Based Information Systems (NBiS), 2011 14th International Conference on*, pages 663–666. IEEE, 2011.

[43] Chris Clifton and Tamir Tassa. On syntactic anonymity and differential privacy. *Transactions on Data Privacy*, 6(2):161–183, 2013.

[44] Graham Cormode. Personal privacy vs population privacy: learning to attack anonymization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1253–1261. ACM, 2011.

[45] Graham Cormode, Divesh Srivastava, Ninghui Li, and Tiancheng Li. Minimizing minimality and maximizing utility: analyzing method-based attacks on anonymized data. *Proceedings of the VLDB Endowment*, 3(1-2):1045–1056, 2010.

[46] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital watermarking and steganography*. Morgan Kaufmann, 2007.

[47] Ingemar J Cox and Matt L Miller. The first 50 years of electronic watermarking. *EURASIP Journal on Advances in Signal Processing*, 2002(2):820936, 2002.

[48] Leucio Antonio Cutillo, Mark Manulis, and Thorsten Strufe. Security and privacy in online social networks. In *Handbook of Social Network Technologies and Applications*, pages 497–522. Springer, 2010.

[49] Leucio Antonio Cutillo, Refik Molva, and Thorsten Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *Communications Magazine*, 47(12):94–101, 2009.

[50] Pradipto Das and D Mandal. Statistical Outlier Detection in Large Multivariate Datasets. *acsu.buffalo.edu*, pages 1–9, 2004.

[51] Emiliano De Cristofaro, Mark Manulis, and Bertram Poettering. Private discovery of common social contacts. In Javier Lopez and Gene Tsudik, editors, *Applied Cryptography and Network Security*, volume 6715 of *Lecture Notes in Computer Science*, pages 147–165. Springer, 2011.

[52] Rinku Dewri, Indrajit Ray, and David Whitley. k-anonymization in the presence of publisher preferences. *Knowledge and Data Engineering, IEEE Transactions on*, 23(11):1678–1690, 2011.

[53] Josep Domingo-Ferrer. Microaggregation: achieving k-anonymity with quasi-optimal data quality. In *Proceeding of 2006 European Conference on Quality in Survey Statistics, Cardiff, UK*, 2006.

[54] Josep Domingo-Ferrer. Rational privacy disclosure in social networks. In *Proceedings of the 7th International Conference on Modeling Decisions for Artificial Intelligence (MDAI), Perpignan, France, October 27-29*, volume 6408, pages 255–265. Springer, 2010.

[55] Josep Domingo-Ferrer. Coprivacy: towards a theory of sustainable privacy. In *Privacy in Statistical Databases*, pages 258–268. Springer, 2011.

[56] Josep Domingo-Ferrer. Rational enforcement of digital oblivion. In *Proceedings of the 4th International Workshop on Privacy and Anonymity in the Information Society*, PAIS '11, pages 2:1–2:8. ACM, 2011.

[57] Josep Domingo-Ferrer and Úrsula González-Nicolás. Hybrid microdata using microaggregation. *Information Sciences*, 180(15):2834–2844, 2010.

[58] Josep Domingo-Ferrer and Josep Maria Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *Knowledge and Data Engineering, IEEE Transactions on*, 14(1):189–201, 2002.

[59] Josep Domingo-Ferrer, Agusti Solanas, and Antoni Martinez-Balleste. Privacy in statistical databases: k-anonymity through microaggregation. In *GrC*, pages 774–777, 2006.

[60] Josep Domingo-Ferrer and Vicenc Torra. Ordinal, continuous and heterogeneous k-anonymity through microaggregation. *Data Mining and Knowledge Discovery*, 11:195–212(18), September 2005.

[61] Judith S Donath. Identity and deception in the virtual community. In Peter Kollock and Marc Smith, editors, *Communities in cyberspace*, pages 29–59. Routledge, 1999.

[62] JohnR. Douceur. The sybil attack. In Peter Druschel, Frans Kaashoek, and Antony Rowstron, editors, *Peer-to-Peer Systems*, volume 2429 of *Lecture Notes in Computer Science*, pages 251–260. Springer Berlin Heidelberg, 2002.

[63] Cynthia Dwork. Differential privacy. In *Automata, languages and programming*, pages 1–12. Springer, 2006.

[64] Manuel Egele, Christopher Kruegel, Engin Kirda, and Giovanni Vigna. Pios: Detecting privacy leaks in ios applications. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th February 2011*. The Internet Society, 2011.

[65] Khaled El Emam, David Buckeridge, Robyn Tamblyn, Angelica Neisa, Elizabeth Jonker, and Aman Verma. The re-identification risk of canadians from longitudinal demographics. *BMC medical informatics and decision making*, 11(1):46, 2011.

[66] Khaled El Emam, Fida Kamal Dankar, Romeo Issa, Elizabeth Jonker, Daniel Amyot, Elise Cogo, Jean-Pierre Corriveau, Mark Walker, Sadrul Chowdhury, Regis Vaillancourt, et al. A globally optimal k-anonymity method for the de-identification of health data. *Journal of the American Medical Informatics Association*, 16(5):670–682, 2009.

[67] Khaled El Emam, Elizabeth Jonker, Luk Arbuckle, and Bradley Malin. A systematic review of re-identification attacks on health data. *PLoS One*, 6(12):e28071, 2011.

[68] Khaled El Emam and Patricia Kosseim. Privacy interests in prescription data, part 2: patient privacy. *Security & Privacy, IEEE*, 7(2):75–78, 2009.

[69] Mark Elliot and Kingsley Purdam. The evaluation of risk from identification attempts, casc project deliverable 5d3, 2003.

[70] William Enck, Peter Gilbert, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. Taintdroid: an information flow tracking system for real-time privacy monitoring on smartphones. *Communications of the ACM*, 57(3):99–106, 2014.

[71] William Enck, Damien Octeau, Patrick McDaniel, and Swarat Chaudhuri. A study of android application security. In *Proceedings of the 20th USENIX Conference on Security*, SEC'11, pages 21–21, Berkeley, CA, USA, 2011. USENIX Association.

[72] European Commission. European union data protection directive 95/46/ec.

[73] Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 211–222. ACM, 2003.

[74] Sascha Fahl, Marian Harbach, Thomas Muders, Lars Baumgärtner, Bernd Freisleben, and Matthew Smith. Why eve and mallory love android: An analysis of android ssl (in)security. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 50–61, New York, NY, USA, 2012. ACM.

[75] Liyue Fan and Hongxia Jin. A practical framework for privacy-preserving data analytics. In *Proceedings of the 24th International Conference on World Wide Web*, pages 311–321. ACM, 2015.

[76] Csilla Farkas and Sushil Jajodia. The inference problem: a survey. *ACM SIGKDD Explorations Newsletter*, 4(2):6–11, 2002.

[77] Federal Court of Canada. Federal Court: Canada (2007) Mike Gordon v. the Minister of Health and the Privacy Commissioner of Canada: Memorandum of Fact and Law of the Privacy Commissioner of Canada. Federal Court., 2007.

[78] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.

[79] Emilio Ferrara, Wen-Qiang Wang, Onur Varol, Alessandro Flammini, and Aram Galstyan. Predicting online extremism, content adopters, and interaction reciprocity. In *International conference on social informatics*, pages 22–39. Springer, 2016.

[80] M. Fire, R. Goldschmidt, and Y. Elovici. Online social networks: Threats and solutions. *IEEE Communications Surveys Tutorials*, 16(4):2019–2036, Fourthquarter 2014.

[81] Dan Frankowski, Dan Cosley, Shilad Sen, Loren Terveen, and John Riedl. You are what you say: privacy risks of public mentions. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 565–572. ACM, 2006.

[82] Carlos Freitas, Fabricio Benevenuto, Saptarshi Ghosh, and Adriano Veloso. Reverse engineering socialbot infiltration strategies in Twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 25–32. ACM, 2015.

[83] Ada Wai-Chee Fu, Ke Wang, Raymond Chi-Wing Wong, Jia Wang, and Minhao Jiang. Small sum privacy and large sum utility in data publishing. *Journal of biomedical informatics*, 50:20–31, 2014.

[84] Benjamin CM Fung, Thomas Trojer, Patrick CK Hung, Li Xiong, Khalil Al-Hussaeni, and Rachida Dssouli. Service-oriented architecture for high-dimensional private data mashup. *IEEE Transactions on Services Computing*, 5(3):373–386, 2012.

[85] Benjamin CM Fung, Ke Wang, and Philip S Yu. Top-down specialization for information and privacy preservation. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 205–216. IEEE, 2005.

[86] Benjamin CM Fung, Ke Wang, and Philip S Yu. Anonymizing classification data for privacy preservation. *Knowledge and Data Engineering, IEEE Transactions on*, 19(5):711–725, 2007.

[87] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith. Composition attacks and auxiliary information in data privacy. In *KDD*, pages 265–273, 2008.

[88] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen. Security issues in online social networks. *IEEE Internet Computing*, 15(4):56–63, July 2011.

[89] Gabriel Ghinita, Panagiotis Karras, Panos Kalnis, and Nikos Mamoulis. Fast Data Anonymization with Low Information Loss. In *VLDB*, 2007.

[90] Gabriel Ghinita, Yufei Tao, and Panos Kalnis. On the Anonymization of Sparse High-Dimensional Data. In *ICDE*, 2008.

[91] Bobi Gilburd, Assaf Schuster, and Ran Wolff. k-ttp: A new privacy model for large-scale distributed environments. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 563–568, New York, NY, USA, 2004. ACM.

[92] Aristides Gionis, Arnon Mazza, and Tamir Tassa. k-anonymization revisited. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 744–753. IEEE, 2008.

[93] Aris Gkoulalas-Divanis, Grigorios Loukides, and Jimeng Sun. Publishing data from electronic health records while preserving privacy: A survey of algorithms. *Journal of Biomedical Informatics*, 50:4 – 19, 2014. Special Issue on Informatics Methods in Medical Privacy.

[94] Olga Gkountouna, Sotiris Angeli, Athanasios Zigomitros, Manolis Terrovitis, and Yannis Vassiliou. k^m-anonymity for continuous data using dynamic hierarchies. In Josep Domingo-Ferrer, editor, *Privacy in Statistical Databases*, volume 8744 of *Lecture Notes in Computer Science*, pages 156–169. Springer International Publishing, 2014.

[95] Philippe Golle. Revisiting the uniqueness of simple demographics in the us population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 77–80. ACM, 2006.

[96] Slawomir Goryczka, Li Xiong, and Benjamin CM Fung. m-privacy for collaborative data publishing. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2011 7th International Conference on*, pages 1–10. IEEE, 2011.

[97] Michael C Grace, Yajin Zhou, Zhi Wang, and Xuxian Jiang. Systematic detection of capability leaks in stock android smartphones. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8*, 2012.

[98] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, WPES '05, pages 71–80, New York, NY, USA, 2005. ACM.

[99] Saikat Guha, Kevin Tang, and Paul Francis. Noyb: Privacy in online social networks. In *Proceedings of the 1st workshop on Online Social Networks*, volume 1, pages 49–54. ACM, 2008.

[100] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*, pages 729–736. ACM, 2013.

[101] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. In *SIGMOD*, pages 1–12, 2000.

[102] Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava. Anonymizing social networks. *Computer Science Department Faculty Publication Series*, page 180, 2007.

[103] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys (CSUR)*, 42(1):1–31, December 2009.

[104] G. Hogben. Security issues and recommendations for online social networks. *ENISA Position Paper*, 1, 2007.

[105] Chia Lung Albert Hsieh, Justin Zhan, Deniel Zeng, and Feiyue Wang. Preserving privacy in joining recommender systems. In *Information Security and Assurance, 2008. ISA 2008. International Conference on*, pages 561–566, April 2008.

[106] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 103–112. ACM, 2011.

[107] Markus Huber, Martin Mulazzani, Edgar Weippl, Gerhard Kitzler, and Sigrun Goluch. Exploiting social networking sites for spam. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, CCS '10, pages 693–695. ACM, 2010.

[108] Markus Huber, Martin Mulazzani, Edgar Weippl, Gerhard Kitzler, and Sigrun Goluch. Friend-in-the-middle attacks: Exploiting social networking sites for spam. *Internet Computing*, 15(3):28–34, 2011.

[109] Eric Hughes. A cypherpunk's manifesto. http://www.activism.net/cypherpunk/manifesto.html, 1993.

[110] IBM. 10 key marketing trends for 2017. https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN, 2017.

[111] internet.org. A Focus on Efficiency white paper from Facebook, Ericsson and Qualcomm. Sep 2014.

[112] Vijay S. Iyengar. Transforming Data to Satisfy Privacy Constraints. In *KDD*, pages 279–288, 2002.

[113] Tom N Jagatic, Nathaniel A Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[114] Sonia Jahid, Prateek Mittal, and Nikita Borisov. Easier: encryption-based access control in social networks with efficient revocation. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pages 411–415. ACM, 2011.

[115] Zhanglong Ji, Zachary C Lipton, and Charles Elkan. Differential privacy and machine learning: a survey and review. *CoRR*, abs/1412.7584, 2014.

[116] Xin Jin, Nan Zhang, and Gautam Das. Algorithm-safe privacy-preserving data publishing. In *Proceedings of the 13th International Conference on Extending Database Technology*, pages 633–644. ACM, 2010.

[117] Xin Jin, Nan Zhang, and Gautam Das. Asap: Eliminating algorithm-based disclosure in privacy-preserving data publishing. *Information Systems*, 36(5):859–880, 2011.

[118] P. Joshi and C. C. J. Kuo. Security and privacy in online social networks: A survey. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6, July 2011.

[119] C Kaleli and H Polat. P2P collaborative filtering with privacy. *Turkish Journal of Electric Electrical Engineering and Compute Science*, 18(1):101–116, 2010.

[120] Miltiadis Kandias, Lilian Mitrou, Vasilis Stavrou, and Dimitris Gritzalis. Which side are you on? A new Panopticon vs. privacy. In *Proceedings of the 10th International Conference on Security and Cryptography (SECRYPT)*, pages 98–110, 2013.

[121] Krishna Kant, Ravishankar Iyer, and Prasant Mohapatra. Architectural impact of secure socket layer on internet servers. In *Proceedings of the International Conference on Computer Design*, pages 7–14. IEEE, 2000.

[122] Dimitrios Karapiperis and Vassilios S Verykios. An lsh-based blocking approach with a homomorphic matching technique for privacy-preserving record linkage. *Knowledge and Data Engineering, IEEE Transactions on*, 27(4):909–921, 2015.

[123] Daniel Kifer. Attacks on privacy and definetti's theorem. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pages 127–138. ACM, 2009.

[124] Daniel Kifer and Johannes Gehrke. Injecting Utility into Anonymized Datasets. In *SIGMOD*, pages 217–228, 2006.

[125] Florian Kohlmayer, Fabian Prasser, Claudia Eckert, Alfons Kemper, and Klaus A Kuhn. Flash: efficient, stable and optimal k-anonymity. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 708–717. IEEE, 2012.

[126] Florian Kohlmayer, Fabian Prasser, Claudia Eckert, and Klaus A Kuhn. A flexible approach to distributed data anonymization. *Journal of biomedical informatics*, 50:62–76, 2014.

[127] B Könings, David Piendl, Florian Schaub, and Michael Weber. Privacyjudge: Effective privacy controls for online published information. In *3rd international conference on Privacy, security, risk and trust (PASSAT), and 3rd International Conference on Social Computing (SocialCom)*, pages 935–941. IEEE, 2011.

[128] Matthijs R. Koot, Guido van't Noordende, and Cees de Laat. A study on the re-identifiability of dutch citizens. *HotPETs 2010 proceedings*, 2010.

[129] B. Krishnamurthy. Privacy and online social networks: can colorless green ideas sleep furiously? *IEEE Security Privacy*, 11(3):14–20, May 2013.

[130] Peter Kwok, Michael Davern, Elizabeth Hair, and Deborah Lafky. Harder than you think: a case study of re-identification risk of hipaa-compliant records. *Chicago: NORC at The University of Chicago. Abstract*, 302255, 2011.

[131] Deborah Lafky. The safe harbor method of de-identification: An empirical test. *Fourth National HIPAA Summit West*, 2010.

[132] Kristen LeFevre, David J. DeWitt, and Raghu Ramakrishnan. Incognito: Efficient Full-domain k-Anonymity. In *SIGMOD*, pages 49–60, 2005.

[133] Kristen LeFevre, David J. DeWitt, and Raghu Ramakrishnan. Mondrian Multidimensional k-Anonymity. In *ICDE*, 2006.

[134] Jiexing Li, Yufei Tao, and Xiaokui Xiao. Preservation of proximity privacy in publishing numerical sensitive data. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 473–486. ACM, 2008.

[135] Ming Li, Ning Cao, Shucheng Yu, and Wenjing Lou. Findu: Privacy-preserving personal profile matching in mobile social networks. In *Proceedings of INFOCOM*, pages 2435–2443. IEEE, 2011.

[136] Na Li, Nan Zhang, and Sajal Das. Preserving relation privacy in online social network data. *IEEE internet computing*, 15(3):35–42, 2011.

[137] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *ICDE*, pages 106–115, 2007.

[138] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. Closeness: A new privacy measure for data publishing. *Knowledge and Data Engineering, IEEE Transactions on*, 22(7):943–956, 2010.

[139] Tiancheng Li and Ninghui Li. On the tradeoff between privacy and utility in data publishing. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 517–526. ACM, 2009.

[140] Tiancheng Li, Ninghui Li, Jian Zhang, and Ian Molloy. Slicing: A new approach for privacy preserving data publishing. *Knowledge and Data Engineering, IEEE Transactions on*, 24(3):561–574, 2012.

[141] Zhen Li and Xiaojun Ye. Privacy protection on multiple sensitive attributes. In *Information and Communications Security*, pages 141–152. Springer, 2007.

[142] Hongyu Liang and Hao Yuan. On the complexity of t-closeness anonymization and related problems. In *Database Systems for Advanced Applications*, pages 331–345. Springer, 2013.

[143] Chi Lin, Zihao Song, Houbing Song, Yanhong Zhou, Yi Wang, and Guowei Wu. Differential privacy preserving in big data analytics for connected health. *Journal of medical systems*, 40(4):1–9, 2016.

[144] Jinfei Liu, Jun Luo, and Joshua Zhexue Huang. Rating: Privacy preservation for multiple attributes with different sensitivity requirements. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 666–673. IEEE, 2011.

[145] Junqiang Liu and Ke Wang. On optimal anonymization for l^+-diversity. In *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*, pages 213–224. IEEE, 2010.

[146] Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6:1–6:30, December 2010.

[147] G. Loukides and A. Gkoulalas-Divanis. Utility-aware anonymization of diagnosis codes. *Biomedical and Health Informatics, IEEE Journal of*, 17(1):60–70, Jan 2013.

[148] Grigorios Loukides, Aris Gkoulalas-Divanis, and Bradley Malin. Anonymization of electronic medical records for validating genome-wide association studies. *Proceedings of the National Academy of Sciences*, 107(17):7898–7903, 2010.

[149] Grigorios Loukides and Jianhua Shao. Preventing range disclosure in k-anonymised data. *Expert Systems with Applications*, 38(4):4559–4574, 2011.

[150] Matthew M Lucas and Nikita Borisov. Flybynight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM workshop on Privacy in the Electronic Society*, pages 1–8. ACM, 2008.

[151] Wanying Luo, Qi Xie, and Urs Hengartner. Facecloak: An architecture for user privacy on social networking sites. In *International Conference on Computational Science and Engineering (CSE'09)*, volume 3, pages 26–33. IEEE, Aug 2009.

[152] N. Lykousas, V. Gómez, and C. Patsakis. Adult content in Social Live Streaming Services: Characterizing deviant users and relationships. *ArXiv e-prints*, June 2018.

[153] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[154] Nidhi Maheshwarkar, Kshitij Pathak, and Narendra S Choudhari. K-anonymity model for multiple sensitive attributes. *Special Issue of International Journal of Computer Applications (0975–8887) on Optimization and On-chip Communication*, 2012.

[155] Majid Bashir Malik, M Asger Ghazi, and Rashid Ali. Privacy preserving data mining techniques: current scenario and future prospects. In *Computer and Communication Technology (ICCCT), 2012 Third International Conference on*, pages 26–32. IEEE, 2012.

[156] Bradley Malin. Re-identification of familial database records. In *AMIA Annual Symposium Proceedings*, volume 2006, page 524. American Medical Informatics Association, 2006.

[157] Josep Maria Mateo-Sanz, Antoni Martínez-Ballesté, and Josep Domingo-Ferrer. Fast generation of accurate synthetic microdata. In *Privacy in Statistical Databases*, pages 298–306. Springer, 2004.

[158] Viktor Mayer-Schönberger. *Delete: the virtue of forgetting in the digital age*. Princeton University Press, 2011.

[159] Pooya Mehregan and Philip WL Fong. Policy negotiation for co-owned resources in relationship-based access control. In *Proceedings of the 21st ACM on Symposium on Access Control Models and Technologies*, pages 125–136. ACM, 2016.

[160] Adam Meyerson and Ryan Williams. On the Complexity of Optimal K-anonymity. In *PODS*, pages 223–228, 2004.

[161] Fred Mintzer and Gordon W Braudaway. If one watermark is good, are more better? In *Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2067–2069. IEEE, 1999.

[162] Noman Mohammed, Benjamin Fung, Patrick CK Hung, and Cheuk-Kwong Lee. Centralized and distributed anonymization for high-dimensional healthcare data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(4):18, 2010.

[163] Noman Mohammed, Benjamin C Fung, and Mourad Debbabi. Anonymity meets game theory: secure data integration with malicious participants. *The VLDB Journal–The International Journal on Very Large Data Bases*, 20(4):567–588, 2011.

[164] Tayana Morkel. The osn-tagging scheme: Recoverable steganography for online social networks. In *Next Generation Computing Applications (NextComp), 2017 1st International Conference on*, pages 11–16. IEEE, 2017.

[165] Krishnamurty Muralidhar and Rathindra Sarathy. Generating sufficiency-based non-synthetic perturbed data. *Transactions on Data Privacy*, 1(1):17–33, 2008.

[166] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *In SP '08: IEEE Symposium on Security and Privacy*, pages 111–125, 2008.

[167] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 173–187. IEEE, 2009.

[168] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of personally identifiable information. *Communications of the ACM*, 53(6):24–26, 2010.

[169] Arvind Narayanan, Narendran Thiagarajan, Mugdha Lakhani, Michael Hamburg, and Dan Boneh. Location privacy via private proximity testing. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th February 2011*. The Internet Society, 2011.

[170] M. Ercan Nergiz and Chris Clifton. Thoughts on k-anonymization. *Data and Knowledge Engineering*, 63(3):622–645, 2007.

[171] Mehmet Ercan Nergiz, Maurizio Atzori, and Chris Clifton. Hiding the presence of individuals from shared databases. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 665–676. ACM, 2007.

[172] Mehmet Ercan Nergiz and Chris Clifton. δ-presence without complete world knowledge. *Knowledge and Data Engineering, IEEE Transactions on*, 22(6):868–883, 2010.

[173] Mehmet Ercan Nergiz, Chris Clifton, and Ahmet Erhan Nergiz. Multirelational k-anonymity. *Knowledge and Data Engineering, IEEE Transactions on*, 21(8):1104–1117, 2009.

[174] Mehmet Ercan Nergiz and Muhammed Zahit Gök. Hybrid k-anonymity. *Computers & Security*, 44:51–63, 2014.

[175] Michel Netter, Moritz Riesner, Michael Weber, and Gunther Pernul. Privacy settings in online social networks–preferences, perception, and reality. In *46th International Conference on System Sciences (HICSS)*, pages 3219–3228. IEEE, 2013.

[176] Jianxia Ning, Indrajeet Singh, Harsha V Madhyastha, Srikanth V Krishnamurthy, Guohong Cao, and Prasant Mohapatra. Secret message sharing using online social media. In *Communications and Network Security (CNS), 2014 IEEE Conference on*, pages 319–327. IEEE, 2014.

[177] Salvador Ochoa, Jamie Rasmussen, Christine Robson, and Michael Salib. Reidentification of individuals in chicago's homicide database: A technical and legal study. *Massachusetts Institute of Technology*, 2001.

[178] Anna Oganian and Josep Domingo-Ferrer. Hybrid microdata via model-based clustering. In *Privacy in Statistical Databases*, pages 103–115. Springer, 2012.

[179] E. Palomar, L. González-Manzano, A. Alcaide, and Á. Galán. Implementing a privacy-enhanced attribute-based credential system for online social networks with co-ownership management. *IET Information Security*, 10(2):60–68, 2016.

[180] A. Papageorgiou, M. Strigkos, E. Politou, E. Alepis, A. Solanas, and C. Patsakis. Security and privacy analysis of mobile health applications: The alarming state of practice. *IEEE Access*, PP(99):1–1, 2018.

[181] A. Papageorgiou, A. Zigomitros, and C. Patsakis. Personalising and crowdsourcing stress management in urban environments via s-health. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–4, July 2015.

[182] Namje Park. Secure data access control scheme using type-based re-encryption in cloud environment. In *Semantic Methods for Knowledge Management and Communication*, pages 319–327. Springer, 2011.

[183] Constantinos Patsakis. Encrypt to forget. In *XII Spanish Meeting on Cryptology and Information Security (RECSI 2012)*, 2012.

[184] Constantinos Patsakis, Alexandros Asthenidis, and Abraham Chatzidimitriou. Social networks as an attack platform: Facebook case study. In *8th International Conference on Networks*, pages 245–247. IEEE, 2009.

[185] Constantinos Patsakis, Michael Clear, Paul Laird, Athanasios Zigomitros, and Mélanie Bouroche. Privacy-aware large-scale virological and epidemiological data monitoring. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014* [4], pages 78–81.

[186] Constantinos Patsakis and Agusti Solanas. Privacy as a product: A case study in the m-health sector. In *4th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–6. IEEE, July 2013.

[187] Constantinos Patsakis and Agusti Solanas. Trading privacy in the cloud: A fairer way to share private information. In *10th IEEE International Conference on e-Business Engineering (ICEBE)*, pages 413–418. IEEE, 2013.

[188] Constantinos Patsakis, Athanasios Zigomitros, Achilleas Papageorgiou, and Edgar Galván López. Distributing privacy policies over multimedia content across multiple online social networks. *Computer Networks*, 75:531–543, 2014.

[189] Constantinos Patsakis, Athanasios. Zigomitros, Achilleas Papageorgiou, and Agusti Solanas. Privacy and security for multimedia content shared on osns: Issues and countermeasures. *The Computer Journal*, 2014. Accepted for publication.

[190] Constantinos Patsakis, Athanasios Zigomitros, Achilleas Papageorgiou, and Agusti Solanas. Privacy and security for multimedia content shared on osns: Issues and countermeasures. *Comput. J.*, 58(4):518–535, 2015.

[191] Constantinos Patsakis, Athanasios Zigomitros, and Agusti Solanas. Analysis of privacy and security exposure in mobile dating applications. In Selma Boumerdassi, Samia Bouzefrane, and Éric Renault, editors, *Mobile, Secure, and Programmable Networking - First International Conference, MSPN 2015, Paris, France, June 15-17, 2015, Selected Papers*, volume 9395 of *Lecture Notes in Computer Science*, pages 151–162. Springer, 2015.

[192] Constantinos Patsakis, Athanasios Zigomitros, and Agusti Solanas. Privacy-aware genome mining: Server-assisted protocols for private set intersection and pattern matching. In Caetano Traina Jr., Pedro Pereira Rodrigues, Bridget Kane, Paulo Mazzoncini de Azevedo Marques, and Agma Juci Machado Traina, editors, *28th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2015, Sao Carlos, Brazil, June 22-25, 2015*, pages 276–279. IEEE, 2015.

[193] Jessie Pease and Julien Freudiger. Engineering privacy for your users. http://devstreaming.apple.com/videos/wwdc/2016/709tvxadw201avg5v7n/709/709_engineering_privacy_for_your_users.pdf, 2016.

[194] Jian Pei, Jian Xu, Zhibin Wang, Wei Wang, and Ke Wang. Maintaining k-anonymity against incremental updates. In *Scientific and Statistical Database Management, 2007. SSBDM'07. 19th International Conference on*, pages 5–5. IEEE, 2007.

[195] Fabien AP Petitcolas. Watermarking schemes evaluation. *IEEE signal processing magazine*, 17(5):58–64, 2000.

[196] Fabien AP Petitcolas and Stefan Katzenbeisser. *Information Hiding Techniques for Steganography and Digital Watermarking (Artech House computer security series)*. Artech House, 2000.

[197] Huseyin Polat and Wenliang Du. Privacy-preserving top-n recommendation on distributed data. *Journal of the Association for Information Science and Technology*, 59(7):1093–1108, 2008.

[198] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. Forgetting personal data and revoking consent under the gdpr: Challenges and proposed solutions. *Journal of Cybersecurity*, 2018.

[199] Eugenia Politou, Alexandra Michota, Efthimios Alepis, Matthias Pocs, and Constantinos Patsakis. Backups and the right to be forgotten in the gdpr: An uneasy relationship. *Computer Law and Security Review*, (6), 2018.

[200] Fabian Prasser, Raffael Bild, Johanna Eicher, Helmut Spengler, Florian Kohlmayer, and Klaus A. Kuhn. Lightning: Utility-driven anonymization of high-dimensional data. *Trans. Data Privacy*, 9(2):161–185, August 2016.

[201] Guojun Qin, Constantinos Patsakis, and Mélanie Bouroche. Playing hide and seek with mobile dating applications. In Nora Cuppens-Boulahia, Frédéric Cuppens, Sushil Jajodia, Anas Abou El Kalam, and Thierry Sans, editors, *ICT Systems Security and Privacy Protection*, volume 428 of *IFIP Advances in Information and Communication Technology*, pages 185–196. Springer Berlin Heidelberg, 2014.

[202] Ariel Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. In *Proceedings of the 4th Symposium on Usable Privacy and Security*, pages 13–23. ACM, 2008.

[203] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Record*, pages 1–20, 2000.

[204] Jerome P Reiter. Inference for partially synthetic, public use microdata sets. *Survey Methodology*, 29(2):181–188, 2003.

[205] Daniele Riboni and Claudio Bettini. Incremental release of differentially-private check-in data. *Pervasive Mob. Comput.*, 16(PB):220–238, January 2015.

[206] Moritz Riesner, Michael Netter, and Günther Pernul. Analyzing settings for social identity management on social networking sites: Classification, current state, and proposed developments. *Information Security Technical Report*, 17(4):185–198, 2013.

[207] D. Rosenblum. What anyone can know: The privacy risks of social networking sites. *IEEE Security Privacy*, 5(3):40–49, May 2007.

[208] Donald B Rubin. Statistical disclosure limitation. *Journal of official Statistics*, 9(2):461–468, 1993.

[209] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[210] Patrick Ruch, Robert H Baud, Anne-Marie Rassinoux, Pierrette Bouillon, and Gilbert Robert. Medical document anonymization with a semantic lexicon. In *Proceedings of the AMIA Symposium*, page 729. American Medical Informatics Association, 2000.

[211] P. Samarati. Protecting Respondents' Identities in Microdata Release. *IEEE TKDE*, 13(6):1010–1027, 2001.

[212] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, Technical report, SRI International, 1998.

[213] Lalitha Sankar, S Raj Rajagopalan, and H Vincent Poor. Utility-privacy tradeoffs in databases: An information-theoretic approach. *Information Forensics and Security, IEEE Transactions on*, 8(6):838–852, 2013.

[214] Pratik Savla and Lorenzo D Martino. Content analysis of privacy policies for health social networks. In *Proceedings of the 2012 IEEE International Symposium on Policies for Distributed Systems and Networks*, pages 94–101. IEEE, 2012.

[215] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web*, pages 745–750. International World Wide Web Conferences Steering Committee, 2016.

[216] Nicholas Paul Sheppard, Reihaneh Safavi-Naini, and Philip Ogunbona. On multiple watermarking. In *Proceedings of the 2001 workshop on Multimedia and Security: New challenges*, pages 3–6. ACM, 2001.

[217] Changgui Shi and Bharat Bhargava. A fast MPEG video encryption algorithm. In *Proceedings of the 6th ACM International Conference on Multimedia*, pages 81–88. ACM, 1998.

[218] Erez Shmueli and Tamir Tassa. Privacy by diversity in sequential releases of databases. *Inf. Sci.*, 298(C):344–372, March 2015.

[219] Erez Shmueli, Tamir Tassa, Raz Wasserstein, Bracha Shapira, and Lior Rokach. Limiting disclosure of sensitive data in sequential releases of databases. *Information Sciences*, 191:98–127, 2012.

[220] Garima Singh and Vijay Kumar. An efficient clustering and distance based approach for outlier detection. *International Journal of Computer Trends and Technology (IJCTT)*, 4(7), 2013.

[221] Andrzej Skowron and Cecylia Rauszer. The discernibility matrices and functions in information systems. In *Intelligent Decision Support*, pages 331–362. Springer, 1992.

[222] Agusti Solanas, Antoni Martinez-Balleste, and Josep M Mateo-Sanz. Distributed architecture with double-phase microaggregation for the private sharing of biomedical data in mobile health. *Information Forensics and Security, IEEE Transactions on*, 8(6):901–910, 2013.

[223] Jordi Soria-Comas and Josep Domingo-Ferrer. Co-utile collaborative anonymization of microdata. In *Modeling Decisions for Artificial Intelligence*, pages 192–206. Springer, 2015.

[224] Jordi Soria-Comas, Josep Domingo-Ferrer, David Sánchez, and Sergio Martínez. t-closeness through microaggregation: Strict privacy with enhanced utility preservation. *IEEE Trans. Knowl. Data Eng.*, 27(11):3098–3110, 2015.

[225] Anna C Squicciarini, Mohamed Shehab, and Joshua Wede. Privacy policies for shared content in social network sites. *The VLDB Journal–The International Journal on Very Large Data Bases*, 19(6):777–796, 2010.

[226] Anna C Squicciarini, Heng Xu, and Xiaolong Luke Zhang. Cope: Enabling collaborative privacy management in online social networks. *Journal of the Association for Information Science and Technology*, 62(3):521–534, 2011.

[227] Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. Collective privacy management in social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 521–530, 2009.

[228] Bowonsak Srisungsittisunti and Juggapong Natwichai. An incremental privacy-preservation algorithm for the (k, e)-anonymous model. *Computers & Electrical Engineering*, 41:126 – 141, 2015.

[229] Klara Stokes and Niklas Carlsson. A peer-to-peer agent community for digital oblivion in online social networks. In *11th Annual International Conference on Privacy, Security and Trust (PST)*, pages 103–110. IEEE, 2013.

[230] Klara Stokes and Vicenç Torra. n-confusion: a generalization of k-anonymity. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 211–215. ACM, 2012.

[231] Katherine Strater and Heather Richter. Examining privacy and disclosure in a social networking community. pages 157–158, 2007.

[232] Xiaoxun Sun, Lili Sun, and Hua Wang. Extended k-anonymity models against sensitive attribute disclosure. *Computer Communications*, 34(4):526–535, 2011.

[233] Supreme Court of the State of Illinois. The Supreme Court of the State of Illinois (2006) Southern Illinoisan vs. The Illinois Department of Public Health, 2006. docket no. 98712.

[234] Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: learning lip sync from audio. *ACM Transactions on Graphics (TOG)*, 36(4):95, 2017.

[235] Latanya Sweeney. Datafly: A system for providing anonymity in medical data. In *Proc. of the International Conference on Database Security*, pages 356–381, 1998.

[236] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, pages 1–34, 2000.

[237] Latanya Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):571–588, 2002.

[238] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[239] Latanya Sweeney. Only you, your doctor, and many others may know. *Technology Science*, 2015.

[240] Latanya Sweeney and Ji Su Yoo. De-anonymizing south korean resident registration numbers shared in prescription data. *Technology Science*, 2015.

[241] György Szarvas, Richárd Farkas, and Róbert Busa-Fekete. State-of-the-art anonymization of medical records using an iterative machine learning framework. *Journal of the American Medical Informatics Association*, 14(5):574–580, 2007.

[242] Krzysztof Szczypiorski. Steghash: New method for information hiding in open social networks. *International Journal of Electronics and Telecommunications*, 62(4):347–352, 2016.

[243] Nilothpal Talukder, Mourad Ouzzani, Ahmed K Elmagarmid, Hazem Elmeleegy, and Mohamed Yakout. Privometer: Privacy protection in social networks. In *IEEE 26th International Conference on Data Engineering Workshops (ICDEW)*, pages 266–269. IEEE, 2010.

[244] Youdong Tao, Yunhai Tong, Shaohua Tan, Shiwei Tang, and Dongqing Yang. T-rotation: Multiple publications of privacy preserving data sequence. In *Advanced Data Mining and Applications*, pages 500–507. Springer, 2008.

[245] Tamir Tassa, Arnon Mazza, and Aristides Gionis. k-concealment: An alternative model of k-type anonymity. *Transactions on Data Privacy*, 5(1):189–222, 2012.

[246] M. Terrovitis, N. Mamoulis, and P. Kalnis. Privacy-preserving Anonymization of Set-valued Data. *PVLDB*, 1(1), 2008.

[247] Manolis Terrovitis, Nikos Mamoulis, John Liagouris, and Spiros Skiadopoulos. Privacy preservation by disassociation. *Proceedings of the VLDB Endowment*, 5(10):944–955, 2012.

[248] Kurt Thomas, Chris Grier, and DavidM. Nicol. unfriendly: Multi-party privacy risks in social networks. In MikhailJ. Atallah and NicholasJ. Hopper, editors, *Privacy Enhancing Technologies*, volume 6205 of *Lecture Notes in Computer Science*, pages 236–252. Springer Berlin Heidelberg, 2010.

[249] Amin Tootoonchian, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. Lockr: better privacy for social networks. In *Proceedings of the 5th International conference on Emerging networking experiments and technologies*, pages 169–180. ACM, 2009.

[250] Vicenç Torra. Microaggregation for categorical variables: A median based approach. In *Privacy in statistical databases*, pages 162–174. Springer, 2004.

[251] Vicenç Torra. Constrained microaggregation: Adding constraints for data editing. *Transactions on Data Privacy*, 1(2):86–104, 2008.

[252] Slim Trabelsi and Hana Bouafif. Abusing social networks with abuse reports - a coalition attack for social networks. In *Proceedings of the 10th International Conference on Security and Cryptography*, 2013.

[253] Traian Marius Truta and Bindu Vinay. Privacy protection: p-sensitive k-anonymity property. In *null*, page 94. IEEE, 2006.

[254] Vassilios S Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin, and Yannis Theodoridis. State-of-the-art in privacy preserving data mining. *ACM Sigmod Record*, 33(1):50–57, 2004.

[255] Vassilios S. Verykios, Ahmed K. Elmagarmid, Elisa Bertino, Yücel Saygin, and Elena Dasseni. Association Rule Hiding. *TKDE*, 16(4), 2004.

[256] Alexandre Viejo, Jordi Castella-Roca, and Guillem Rufián. Preserving the user's privacy in social networking sites. In *Trust, Privacy, and Security in Digital Business*, pages 62–73. Springer, 2013.

[257] Bimal Viswanath, M Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. Towards detecting anomalous user behavior in online social networks. In *Proceedings of the 23rd USENIX Security Symposium (USENIX Security)*, 2014.

[258] Sviatolsav Voloshynovskiy, Shelby Pereira, Thierry Pun, Joachim J Eggers, and Jonathan K Su. Attacks on digital watermarks: classification, estimation based attacks, and benchmarks. *Communications Magazine*, 39(8):118–126, 2001.

[259] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[260] Guan Wang, Zutao Zhu, Wenliang Du, and Zhouxuan Teng. Inference analysis in privacy-preserving data re-publishing. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 1079–1084. IEEE, 2008.

[261] Guojun Wang, Qin Liu, and Jie Wu. Hierarchical attribute-based encryption for fine-grained access control in cloud storage services. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, pages 735–737. ACM, 2010.

[262] Hui (Wendy) Wang and Ruilin Liu. Hiding outliers into crowd: Privacy-preserving data publishing with outliers. *Data & Knowledge Engineering*, 100, Part A:94 – 115, 2015.

[263] Ke Wang and Benjamin Fung. Anonymizing sequential releases. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 414–423. ACM, 2006.

[264] Ke Wang, Benjamin C. M. Fung, and Guozhu Dong. Integrating private databases for data analysis. In *In IEEE ISI*, pages 171–182. Springer Verlag, 2005.

[265] Ke Wang, Benjamin CM Fung, and S Yu Philip. Handicapping attacker's confidence: an alternative to k-anonymization. *Knowledge and Information Systems*, 11(3):345–368, 2007.

[266] Ke Wang, Benjamin CM Fung, and Philip S Yu. Template-based privacy preservation in classification problems. In *Data Mining, Fifth IEEE International Conference on*, pages 8–pp. IEEE, 2005.

[267] Wei Wang, Lei Chen, and Qian Zhang. Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. *Computer Networks*, 88:136–148, 2015.

[268] Peter Wayner. *Disappearing cryptography: information hiding: steganography & watermarking*. Morgan Kaufmann, 2009.

[269] Helena Webb, Pete Burnap, Rob Procter, Omer Rana, Bernd Carsten Stahl, Matthew Williams, William Housley, Adam Edwards, and Marina Jirotka. Digital wildfires: Propagation, verification, regulation, and responsible innovation. *ACM Transactions on Information Systems (TOIS)*, 34(3):15, 2016.

[270] D. Wetherall, D. Choffnes, B. Greenstein, S. Han, P. Hornyack, J. Jung, S. Schechter, and X. Wang. Privacy revelations for web and mobile apps. pages 21–21, 2011.

[271] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. A practical attack to de-anonymize social network users. In *31st IEEE Symposium on Security and Privacy*, pages 223–238. IEEE, 2010.

[272] R. C.-W. Wong, A. W.-C. Fu, K. Wang, and J. Pei. Minimality attack in privacy preserving data publishing. In *VLDB*, pages 543–554, 2007.

[273] Raymond Chi-Wing Wong, Ada Wai-Chee Fu, Ke Wang, Philip S. Yu, and Jian Pei. Can the utility of anonymized data be used for privacy breaches? *ACM Trans. Knowl. Discov. Data*, 5(3):16:1–16:24, August 2011.

[274] Raymond Chi-Wing Wong, Jiuyong Li, Ada Wai-Chee Fu, and Ke Wang. (α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 754–759. ACM, 2006.

[275] Wai Kit Wong, Nikos Mamoulis, and David Wai Lok Cheung. Non-homogeneous generalization in privacy preserving data publishing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 747–758. ACM, 2010.

[276] Yingjie Wu, Xiaowen Ruan, Shangbin Liao, and Xiaodong Wang. P-cover k-anonymity model for protecting multiple sensitive attributes. In *Computer Science and Education (ICCSE), 2010 5th International Conference on*, pages 179–183. IEEE, 2010.

[277] Xiaokui Xiao and Yufei Tao. Anatomy: simple and effective privacy preservation. In *VLDB*, pages 139–150, 2006.

[278] Xiaokui Xiao and Yufei Tao. Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 229–240. ACM, 2006.

[279] Xiaokui Xiao and Yufei Tao. M-invariance: towards privacy preserving re-publication of dynamic datasets. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 689–700. ACM, 2007.

[280] Xiaokui Xiao, Yufei Tao, and Nick Koudas. Transparent anonymization: Thwarting adversaries who know the algorithm. *ACM Transactions on Database Systems (TODS)*, 35(2):8, 2010.

[281] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. Fu. Utility-Based Anonymization Using Local Recoding. In *KDD*, pages 785–790, 2006.

[282] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren. Information security in big data: Privacy and data mining. *IEEE Access*, 2:1149–1176, 2014.

[283] Ibrahim Yakut and Huseyin Polat. Privacy-Preserving Svd-Based Collaborative Filtering on Partitioned Data. *International Journal of Information Technology & Decision Making*, 09(03):473–502, May 2010.

[284] Ibrahim Yakut and Huseyin Polat. Estimating NBC-based recommendations on arbitrarily partitioned data with privacy. *Knowledge-Based Systems*, 36:353–362, December 2012.

[285] Chao Yao, X Sean Wang, and Sushil Jajodia. Checking for k-anonymity violation by views. In *Proceedings of the 31st international conference on Very large data bases*, pages 910–921. VLDB Endowment, 2005.

[286] Yang Ye, Yu Liu, Chi Wang, Dapeng Lv, and Jianhua Feng. Decomposition: Privacy preservation for multiple sensitive attributes. In *Database Systems for Advanced Applications*, pages 486–490. Springer, 2009.

[287] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable, and fine-grained data access control in cloud computing. In *Proceedings of IEEE INFOCOM, 2010*, pages 1–9. IEEE, 2010.

[288] Hessam Zakerzadeh, Charu C. Aggarwal, and Ken Barker. Towards breaking the curse of dimensionality for high-dimensional privacy: An extended version. *CoRR*, abs/1401.1174, 2014.

[289] Hessam Zakerzadeh, Charu C. Aggarwal, and Ken Barker. Managing dimensionality in data privacy anonymization. *Knowledge and Information Systems*, 49(1):341–373, 2016.

[290] C. Zhang, J. Sun, X. Zhu, and Y. Fang. Privacy and security for online social networks: challenges and opportunities. *IEEE Network*, 24(4):13–18, July 2010.

[291] Qing Zhang, Nick Koudas, Divesh Srivastava, and Ting Yu. Aggregate Query Answering on Anonymized Tables. In *ICDE*, pages 116–125, 2007.

[292] Xiaojian Zhang, Xiaofeng Meng, and Rui Chen. *Differentially Private Set-Valued Data Release against Incremental Updates*, pages 392–406. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[293] Li Zhao, Ravi Iyer, Srihari Makineni, and Laxmi Bhuyan. Anatomy and performance of ssl processing. In *IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2005)*, pages 197–206. IEEE, 2005.

[294] Elena Zheleva and Lise Getoor. *Privacy in Social Networks: A Survey*, pages 277–306. Springer US, Boston, MA, 2011.

[295] Bin Zhou, Jian Pei, and WoShun Luk. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *ACM SIGKDD Explorations Newsletter*, 10(2):12–22, 2008.

[296] Yajin Zhou, Xinwen Zhang, Xuxian Jiang, and VincentW. Freeh. Taming information-stealing smartphone applications (on android). In JonathanM. McCune, Boris Balacheff, Adrian Perrig, Ahmad-Reza Sadeghi, Angela Sasse, and Yolanta Beres, editors, *Trust and Trustworthy Computing*, volume 6740 of *Lecture Notes in Computer Science*, pages 93–107. Springer Berlin Heidelberg, 2011.

[297] Athanasios Zigomitros, Achilleas Papageorgiou, and Constantinos Patsakis. Social network content management through watermarking. In Geyong Min, Yulei Wu, Lei (Chris) Liu, Xiaolong Jin, Stephen A. Jarvis, and Ahmed Yassin Al-Dubai, editors, *11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2012, Liverpool, United Kingdom, June 25-27, 2012*, pages 1381–1386. IEEE Computer Society, 2012.

[298] Athanasios Zigomitros, Achilleas Papageorgiou, and Constantinos Patsakis. A practical k-anonymous recommender system. In *Information, Intelligence, Systems & Applications (IISA), 2016 7th International Conference on*, pages 1–4. IEEE, 2016.

[299] Athanasios Zigomitros and Constantinos Patsakis. Cross format embedding of metadata in images using qr codes. In *Intelligent Interactive Multimedia Systems and Services*, volume 11 of *Smart Innovation, Systems and Technologies*, pages 113–121. Springer, 2011.

[300] Athanasios Zigomitros and Constantinos Patsakis. Storing metadata as QR codes in multimedia streams. In Emmanouel Garoufallou and Jane Greenberg, editors, *Metadata and Semantics Research - 7th Research Conference, MTSR 2013, Thessaloniki, Greece, November 19-22, 2013. Proceedings*, volume 390 of *Communications in Computer and Information Science*, pages 152–162. Springer, 2013.

[301] Athanasios Zigomitros, Agusti Solanas, and Constantinos Patsakis. The role of inference in the anonymization of medical records. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014* [4], pages 88–93.