Chapter 14
Privacy-Preserving Technologies

Josep Domingo-Ferrer and Alberto Blanco-Justicia

Abstract This chapter introduces privacy and data protection by design, and reviews privacy-enhancing techniques (PETs). Although privacy by design includes both technical and operational measures, the chapter focuses on the technical measures. First, it enumerates design strategies. Next, it considers privacy-enhancing techniques that directly address the *hide* strategy, but also aid in implementing the *separate*, *control* and *enforce* strategies. Specifically, it addresses PETs for: (1) identification, authentication and anonymity; (2) private communications; (3) privacy-preserving computations; (4) privacy in databases; and (5) discrimination prevention in data mining.

Keywords Anonymisation · Cryptography · Digital signatures · Privacy · Privacy-enhancing techniques · Statistical disclosure control

14.1 Introduction

Applying cybersecurity mechanisms is essential to the protection of digital assets, whether they be personal, industrial or commercial. Current cybersecurity (and safety) measures include the collection of data from several points to detect, and potentially foresee, anomalies that can be attributed to malicious behaviour (e.g. cyberattacks). Collecting these data can, in some cases, encroach on the privacy of citizens. The new General Data Protection Regulation (GDPR)[1] states that the collection and processing of personal data for cybersecurity reasons is legitimate; how-

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

J. Domingo-Ferrer (✉) · A. Blanco-Justicia
Department of Computer Science and Mathematics, CYBERCAT – Center for Cybersecurity Research of Catalonia, UNESCO Chair in Data Privacy, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain
e-mail: josep.domingo@urv.cat; alberto.blanco@urv.cat

M. Christen et al. (eds.), *The Ethics of Cybersecurity*, The International Library of Ethics, Law and Technology 21,
https://doi.org/10.1007/978-3-030-29053-5_14

ever, it is still subject to the rest of requirements of the regulation, such as consent, transparency and adequate protection (see also Chaps. 5 and 10).

This chapter introduces privacy and data protection by design and reviews privacy-enhancing techniques (PETs). Although privacy by design includes both technical and operational measures, we focus here on the technical measures.

Therefore, the analyses within this chapter can empower both cybersecurity service providers and general service providers to design systems that are compliant with the GDPR, in addition to achieving other benefits. For example, while personal data can only be held by a controller for a limited period of time, anonymised data are no longer considered personal data and thus they are outside the scope of GDPR. Hence, anonymised data can be handled much more freely: they can be shared and stored indefinitely, which in particular enables exploratory, collaborative and long-term studies.

14.1.1 Design Strategies

Privacy and data protection by design can be achieved by applying certain design strategies (see also Chap. 2). We next enumerate the eight design strategies introduced by Hoepman (2014).

1. *Minimise.* System designers should ensure that only the minimal necessary personal information is collected.
2. *Hide.* This strategy implies that the confidentiality of collected data is ensured, either by encrypting, pseudonymising or anonymising data in transit or in storage. The rest of this document is dedicated to describing several mechanisms to enforce this design principle.
3. *Separate.* Personal data should be stored and processed in a distributed way.
4. *Aggregate.* Storage of individualised data should be restricted as much as possible and be replaced by aggregates whenever feasible.
5. *Inform.* Respondents should be made aware of what information about them is being collected and processed and for which reasons.
6. *Control.* Respondents should be able to consult, modify and delete the information about them.
7. *Enforce.* Privacy policies should be put in place and enforced.
8. *Demonstrate.* Data controllers ought to document all collection and analysis processes conducted on personal information.

The remaining sections enumerate privacy-enhancing techniques that directly address the *hide* strategy but also aid in implementing the *separate*, *control* and *enforce* strategies. Note that some of the techniques described render data unlinkable to individuals, that is, they turn personal data into data that are no longer personal.

14.2 Identity, Authentication and Anonymity

Identity, authentication and access control are central components of secure systems. It is important that data assets be accessible only to authorised parties. On the one hand, a sound authentication and authorisation infrastructure prevents data breaches. On the other hand, it allows responsibilities to be attributed in case of a breach, which contributes to a transparent data processing environment.

Several methods exist to verify the identity of individuals, that is, to authenticate them. Some of them allow for the authentication of users without disclosing their identity.

14.2.1 Digital Signatures

In paper documents, handwritten signatures guarantee the authenticity of the document, and the signer cannot repudiate it. Moreover, the paper support gives some protection against manipulation: deletions and additions can be detected, at least by an expert. Digital signatures were created in order to guarantee the authenticity and integrity in the case of electronic communications, and to avoid their repudiation. Digital signatures were made possible by the deployment of public-key encryption. In addition, digital signatures and the public key infrastructure can be used to provide authentication of individuals.

If both sender and receiver share some information, an alternative to digital signatures are message authentication codes (MACs). They are based on keyed cryptographic hash functions, and they can be used to guarantee the integrity of the message. MACs are commonly used in the context of symmetric encryption communications, where sender and receiver share a secret key.

Next, we enumerate specific classes of digital signatures that enable authentication while being compatible with some user anonymity.

14.2.1.1 Blind Signatures

Blind signatures (Chaum 1983) are considered particularly useful for electronic payment systems, electronic voting schemes and token-based access control mechanisms; a user may obtain a signature (e.g. a signed coin from a bank) such that the signer does not know the contents of the message and cannot produce further valid signatures.

14.2.1.2 Group Signatures

In a group signature scheme (Chaum and Van Heyst 1991), a set of users, called members of the group, can issue signatures of arbitrary messages on behalf of the group. A verifier can check the validity of the signature using the group public key. The main interest in this kind of signature is that it ensures the privacy of signers against potential verifiers, because a potential verifier cannot distinguish two signers from the same group.

A requirement of group signatures is the support for membership revocation of misbehaving members without the need to update the group public key. To facilitate member revocation, some members, called group managers, are endowed with the capability to revoke membership.

14.2.1.3 Identity-Based Signatures

Identity-based signature schemes, theorised in Shamir (1984) and with the first concrete protocol, based on the Weil pairing, shown in Boneh and Franklin (2001), allow public keys to be arbitrary strings of some length, called identities. These strings are associated with a user and reflect some aspect of her identity, e.g. his email address. The corresponding secret key is then computed by a trusted entity taking as input the user's identity and, possibly, some other secret information, and is sent to the user through some secure channel. Identity-based public key signature schemes offer considerable flexibility in key generation and management.

14.2.1.4 Attribute-Based Signatures

Attribute-based signatures generalise identity-based signatures in that, instead of having the users' identities as credentials, they use properties, or attributes, of the users as the latters' credentials (in the attribute-based setting, the identity is one more attribute of the user). Attribute-based signatures were introduced by Shanqing and Yingpei (2008), inspired by previously existing attribute-based encryption schemes, such as the one in Goyal et al. (2006). In attribute-based signatures (and encryption) schemes, the users receive private key shares associated to their credentials, such as their name, age, country of residence, having or not a driving licence, place of work, etc. Digital signatures are produced with respect to some function of the users' credentials, typically called a *policy*.

For example, a drugstore may accept drug prescriptions only if they are issued by medical doctors or by nurses with a long working experience. In this scenario, prescriptions could be digitally signed under the policy *role = "medical doctor" OR (role = "nurse" AND experience = "10 years")*. The identity of the signer in this case is irrelevant.

14.2.2 Zero-Knowledge Proofs

Zero-knowledge proofs (Ben-Or et al. 1988b) allow a prover to convince a verifying party of the truth of a statement without revealing any information other than the truth of the statement. In particular, if the statement requires the prover to hold some secret information, then the verifier does not learn this information—it is possible to prove knowledge of a secret without revealing the secret itself. Statements that only prove possession of a secret are known as zero-knowledge proofs of knowledge. Proofs can be either interactive or non-interactive depending on whether the parties can communicate during the proof. In general, non-interactive proofs (Blum et al. 1988) are considered more difficult since they cannot use interactive challenge-response protocols and they require the random oracle model or a common reference string between parties. Whereas zero-knowledge proofs can be rather inefficient, non-interactive proof systems built on bilinear groups (Groth and Sahai 2008) are particularly efficient for group-dependent problems where the secrets are group elements or the exponents of a group element. As many useful cryptographic schemes are built using bilinear pairings, particularly functional encryption, such a proof system can be very useful for proving knowledge of a cryptographic secret without revealing it.

Zero-knowledge proofs can be used to authenticate users holding cryptographic devices, such as smartcards, without leaking any information about these users except that they hold a valid card.

14.2.3 Implicit Authentication

In implicit authentication, a server can authenticate users by checking whether their behaviour is compatible or similar enough to their past-recorded behaviour. In this context, the user's behaviour can be modelled as a combination of features such as her browsing history, usual location, keystroke patterns, usually visible cell stations, etc.

In the study of Jakobsson et al. (2009), empirical evidence was given that the features collected from the user's device history are effective to distinguish users and therefore can be used to implicitly authenticate them. The collection of these data, however, may be too privacy-invasive. Proposals such as those by Safa et al. (2014) Domingo-Ferrer et al. (2015) and Blanco-Justicia and Domingo-Ferrer (2018) make use of homomorphic encryption and secure multiparty computation to authenticate users from their past behaviour without forcing them to disclose their profiles.

14.3 Private Communications

This section discusses the protection of communication channels. First, it describes end-to-end encryption, which provides confidentiality of communications. It then introduces anonymous channels. Having discussed mechanisms that allow users to be authenticated without revealing their identities, it is logical to discuss communication channels that do not reveal their address, which is also part of their identity.

14.3.1 End-to-End Encryption

End-to-end encryption refers to the encryption of messages exchanged by two or more parties without the intervention of a centralised server. The centralised server may exist and support the transport of the messages but all this server sees is encrypted content. This behaviour is the opposite of the traditional message exchange protocols, in which the messages are only encrypted while in transit from the parties to the central server or from the central server to the parties.

End-to-end encryption is typically supported by having all participants have a key pair from a public-key encryption scheme. The centralised server, in addition to supporting the exchange of messages, works as a public-key repository, where users can find the public keys of the users to whom they want to send messages. Once a user has obtained another user's public key, she can use this public key to encrypt the messages, which will only be decryptable by the owner of the corresponding private key. A more efficient variant is for users to exchange random session keys for symmetric encryption by enciphering them under their public-private pairs and then encrypting the messages with a symmetric encryption scheme under these random temporal session keys.

14.3.2 Anonymous Channels

Anonymous channels allow users to hide their address (e.g. the IP address) to the service provider they are communicating with. Examples of anonymous channels include mixnets and onion routing.

A mix network or mixnet is a routing protocol in which each of the network nodes shuffles (and re-encrypts) all received messages before sending them to the next node (Chaum 1981). The shuffling process is kept secret by each mix server. Additionally, the sender of the message might successively encrypt the message with each of the mix servers' public keys. If that is the case, each mix server will have to decrypt each of the encryption layers (as if peeling an onion) until the final destination of the message. The ToR network (Dingledine et al. 2004) is an example of this operation.

14.4 Privacy-Preserving Computations

This section describes mechanisms to make computations on data while keeping the data private. The GDPR accepts encryption as a valid protection mechanism if the decryption keys are only available to those entitled to have them. However, most data analyses are incompatible with most encryption procedures: users typically require data in clear form to analyse them.

Nonetheless, the following encryption techniques do allow some computations to be carried out directly on encrypted data, and are usually part of larger systems, such as privacy-preserving data mining.

14.4.1 (Partially) Homomorphic Encryption

Some encryption schemes are homomorphic in nature. Given two ciphertexts encrypting two plaintexts, certain operations can be performed on the ciphertexts such that the result can be decrypted to produce the outcome of applying an operation (not necessarily the same) on the plaintexts themselves. Thus, some computations can be performed on encrypted data. Schemes that exhibit homomorphic properties for a specific operation are known as partially homomorphic encryption schemes. Examples of this class are those in ElGamal (1985) and Paillier (1999). On the other hand, if the set of permissible operations enable arbitrary computations to be performed, then the schemes are referred to as fully homomorphic (Gentry 2009; Gentry et al. 2013). Although fully homomorphic schemes are in principle very powerful, currently available instances also involve very substantial overhead and storage expansion. For that reason, less powerful schemes, known as somewhat homomorphic, are sometimes preferred: under these schemes, the number of operations that can be performed on ciphertext before decryption will no longer succeed is limited.

14.4.2 Multiparty Computation

Secure multiparty computation protocols allow a set of parties to compute a joint function of their inputs in a secure way without requiring a trusted third party. During the execution of the protocol the parties do not learn anything about each other's input except what is implied by the output itself.

A general solution for the secure computation of functions among two players was introduced in Yao (1986). The main idea of these protocols was to describe the function as a circuit, and to compute every gate of the circuit in a secure way. This idea was extended to the multi-partite setting in Goldreich et al. (1987). They showed how to create a secure multiparty computation protocol that allows playing

any game and does not leak any information if the majority of the participants are honest. These protocols are computationally secure. The first unconditionally secure multi-party computation protocols were presented in Ben-Or et al. (1988a) and Chaum et al. (1988). These authors gave protocols to compute any arithmetic function in a secure way when at least two thirds of the parties are honest.

Two of the main open problems in secure multiparty computation are: (i) to relax the assumptions on the behaviour of the players, and (ii) to reduce the computational and communication costs of the protocols for interesting families of functions. It should be observed that, in the general solutions described above, the computational costs of the protocol depend on the size of the circuit defining the function.

The most important properties of secure multiparty computation protocols are privacy and correctness. Another important property is fairness. A protocol is fair if there are no differences between the players when it comes to obtaining the output. That is, a protocol is fair if either everybody receives their output, or no one does.

14.5 Privacy in Databases

An alternative strategy to protect data is to make them no longer linkable to individuals, that is, to anonymise them. Anonymised data are no longer considered personal, and thus the legal restrictions that apply to personal data are lifted. This section describes the state of the art in data anonymisation techniques and models.

14.5.1 Respondent Privacy: Statistical Disclosure Control

Traditionally, national statistical institutes and government agencies have systematically gathered information about individual respondents, either people or companies, with the aim of using it for policymaking and also distributing it for public and private research that may benefit their country. The most detailed way to disseminate this information is by releasing a microdata set, essentially a database table, each of whose records conveys information on a particular respondent. Although these databases may be extremely useful to researchers, it is of fundamental importance that their publication does not compromise the respondents' privacy in the sense of revealing information attributable to specific individuals. Statistical disclosure control (SDC) is the discipline that deals with the inherent trade-off between protecting the privacy of the respondents and ensuring that the protected data are still useful to researchers.

Usually, a microdata set contains a set of attributes that may be classified as identifiers, key attributes (a.k.a. quasi-identifiers), or confidential attributes. Identifiers allow unequivocal identification of individuals. Examples are social security numbers or full names, which need to be removed before publication of the

microdata set. On the other hand, key attributes are those attributes that, in combination, may allow linkage with external information to re-identify (some of) the respondents to whom (some of) the records in the microdata set refer (*identity disclosure*). Examples include job, address, age, gender, height and weight. Last but not least, the microdata set contains confidential attributes with sensitive information on respondents, such as salary, religion, political affiliation or health condition. Beyond protecting against identity disclosure, SDC must prevent intruders from guessing the confidential attribute values of specific respondents (*attribute disclosure*).

Several SDC methods have been proposed in the literature to protect microdata sets (Hundepool et al. 2012). Next, we briefly review the main ones.

14.5.2 Non-perturbative Masking

In SDC, masking refers to the process of obtaining an anonymised data set *X'* by modifying the original *X*. Masking can be perturbative or non-perturbative. In the former approach, the data values of *X* are perturbed to obtain *X'*. In contrast, in non-perturbative masking *X'* is obtained by removing some values and/or by making them more general; yet the information in *X'* is still true, although less detailed; as an example, a value might be replaced by a range containing the original value.

Common non-perturbative methods include:

- *Sampling*. Instead of publishing the whole data set, only a sample of it is released.
- *Generalisation*. The values of the different attributes are recoded in new, more general categories such that the information remains the same, albeit less specific.
- *Top/bottom coding*. In line with the previous method, values above (resp. below) a certain threshold are grouped together into a single category.
- *Local suppression*. If a combination of quasi-identifier values is shared by too few records, it may lead to re-identification. This method relies on replacing certain individual attribute values with missing values, so that the number of records sharing a particular combination of quasi-identifier values becomes larger.

14.5.3 Perturbative Masking

Perturbative masking generates a modified version of the microdata set such that the privacy of the respondents is protected to a certain extent while simultaneously some statistical properties of the data are preserved. Well-known perturbative masking methods include:

– *Noise addition*. This is the most popular method, which consists in adding a noise vector to each record in the data set. The utility preservation depends on the amount and the distribution of the noise.
– *Data swapping*. This technique exchanges the values of the attributes randomly among individual records. Clearly, univariate distributions are preserved, but multivariate distributions may be substantially harmed unless swaps of very different values are ruled out.
– *Microaggregation*. This groups similar records together and releases the average record of each group (Domingo-Ferrer and Mateo-Sanz 2002). The more similar the records in a group, the more data utility is preserved.

14.5.4 Synthetic Microdata Generation

An anonymisation approach alternative to masking is synthetic data generation. That is, instead of modifying the original data set, a simulated data set is generated such that it preserves some properties of the original data set. The main advantage of synthetic data is that no respondent re-identification seems possible since the data are artificial. However, if, by chance, a synthetic record is very close to an original one, the respondent of the latter record will not feel safe when the former record is released. In addition, the utility of synthetic data sets is limited to preserving the statistical properties selected at the time of data synthesis.

Some examples of synthetic generation include methods based on multiple imputation (Rubin 1993) and methods that preserve means and co-variances (Burridge 2003). An effective alternative to the drawbacks of purely synthetic data are hybrid data, which mix original and synthetic data and are therefore more flexible (Domingo-Ferrer and González-Nicolás 2010). Yet another alternative is partially synthetic data, whereby only the most sensitive original data values are replaced by synthetic values.

14.5.5 Privacy Models

For an anonymised data set *X'* to be safe/private enough, it needs to be sufficiently anonymised. The level of anonymisation can be assessed after the generation of *X'* or prior to it.

Ex post methods rely on the analysis of the output data set and, therefore, it is possible to generate a data set that is not safe enough according to a certain criterion; several iterations with increasingly strict privacy parameters and decreasing utility may be needed. The most commonly used *ex post* approach is masking followed by record linkage. Protection is sufficient high only if there is a sufficiently low proportion of masked records that can be linked to the respective original records they come from.

On the other hand, the *ex ante* approach relies on *privacy models* that allow selecting the desired privacy level before producing *X'*. In this way, the output data set is always as private a specified by the model, although it may fail to provide enough utility if the model parameters are too strict.

14.5.5.1 k-Anonymity and Extensions

A well-known privacy model is *k*-Anonymity (Samarati and Sweeney 1998), which requires that each tuple of key-attribute values be shared by at least *k* records in the database. This condition may be achieved through generalisation and suppression mechanisms, and also through microaggregation (Domingo-Ferrer and Torra 2005).

Unfortunately, while this privacy model prevents identity disclosure, it may fail to protect against attribute disclosure. The definition of this privacy model establishes that complete re-identification is unfeasible within a group of records sharing the same tuple of perturbed key-attribute values. However, if the records in the group have the same value (or very similar values) for a confidential attribute, the confidential attribute value of an individual linkable to the group is leaked.

To fix this problem, some extensions of *k*-Anonymity have been proposed, the most popular being *l*-diversity (Machanavajjhala et al. 2006) and *t*-closeness (Li et al. 2007a). The property of *l*-diversity is satisfied if there are at least *l* 'well-represented' values for each confidential attribute in all groups sharing the values of the quasi-identifiers. The property of *t*-closeness is satisfied when the distance between the distribution of each confidential attribute within each group and the whole data set is no more than a threshold *t*.

14.5.5.2 Differential Privacy

Another important privacy model is differential privacy (Dwork 2006). This model was originally defined for queryable databases and consists in perturbing the original query result of a database before outputting it. This may be viewed as equivalent to perturbing the original data and then computing the queries over the modified data. Thus, differential privacy can also be seen as a privacy model for microdata sets.

An ε-differentially private algorithm is one that, when run on two datasets that differ in a single record, performs similarly (up to a power of ε) in both cases. That is, the presence or the absence of any single record does not significantly alter the output of the algorithm. Typically, ε-differential privacy is attained by adding Laplace noise with zero mean and parameter $\Delta(f)/\varepsilon$, where $\Delta(f)$ is the sensitivity of the algorithm (the maximum change in the algorithm output that can be caused by a change in a single record in the absence of noise) and ε is a privacy parameter; the larger ε, the less privacy.

14.5.5.3 Permutation Model for Anonymisation

The permutation model (Domingo-Ferrer and Muralidhar 2016) views all anonymisation methods as being functionally equivalent to a two-step procedure consisting of a permutation step (mapping the original data set to the output of a reverse mapping procedure [Muralidhar et al. 2014]) plus a noise addition step (adding the difference between the reverse-mapped output and the anonymised data set). Since the ranks in the reverse-mapped version and in the anonymised version are the same by construction, the noise added in the second step needs to be small, since otherwise ranks would change. This shows that any anonymisation method basically amounts to permutation.

The most interesting feature, however, is that each subject/respondent can check whether a privacy model called *(d,v)*-permuted privacy is satisfied for his or her original record by the anonymised data set for some d and v of her choice; in plain words, each subject can check whether his or her response has been permuted enough in the anonymised data set. The subject only needs to know his or her original record and the anonymised data set.

14.5.6 Redaction and Sanitisation of Documents

Document redaction consists of removing or blacking out sensitive terms in plain textual documents. Alternatively, when sensitive terms are replaced (instead of removed) by generalisations (e.g. AIDS → disease), the process is more generically referred to as document sanitisation (Bier et al. 2009). Document sanitisation is more desirable than pure redaction, since the former better preserves the utility of the protected output. Moreover, in document redaction, the existence of blacked-out parts in the released document can raise awareness of the document's sensitivity to potential attackers (Bier et al. 2009), whereas sanitisation gives no such clues.

In both cases, two tasks should be performed: (i) the detection of textual terms that may cause disclosure of sensitive information, and (ii) the removal or obfuscation of those entities. Traditionally, the detection of sensitive terms has been tackled in a manual way. This requires a human expert who applies certain standard guidelines that detail the correct procedures to sanitise sensitive entities (National Security Agency 2005). Manual redaction has proven to be quite time-consuming and it does not scale to currently required levels of information outsourcing (Chakaravarthy et al. 2008; Bier et al. 2009).

In recent years, numerous automatic redaction methods have been proposed. Some approaches rely on specific or tailored patterns to detect certain types of information based on their linguistic or structural regularities (e.g. names, addresses and social security numbers) (Sweeney 1996; Tveit et al. 2004; Douglass et al. 2005). Schemes such as Douglass et al. (2005) and Tveit et al. (2004) use more specific patterns to remove sensitive terms from medical records. These patterns are designed according to the HIPAA 'Safe Harbor' rules (Department of Health and Human

Services, USA 1996) that specify eighteen data elements which must be eliminated from clinical data in order to anonymise a clinical text. As an alternative to manually-specified patterns, several authors have proposed using trained classifiers that recognise sensitive entities. Yet others present a tool that focuses on the sanitisation of documents directly linked to certain companies (Cumby and Ghani 2011). The data to be detected include words and phrases that reveal the company the document belongs to.

Abril et al. (2011) propose a general scheme that uses a trained classifier for Named Entity Recognition (NER) (i.e. the Stanford NER [Finkel et al. 2005]) to automatically recognise entities belonging to general categories such as person, organisation and location names. This mechanism suggests generalising sensitive entities instead of removing them from the sanitised document. The goal is to achieve a certain degree of privacy while preserving some of the semantics. Jiang et al. (2009) provide a theoretic measure ('t-plausibility') that guides the sanitisation process in order to balance the trade-off between privacy protection and utility preservation. Their scheme tries to preserve the utility of sanitised documents by generalising terms based on general-purpose ontology/taxonomy. Finally, Sánchez et al. (2013) present a system that relies on information theory to quantify the amount of information conveyed by each term of the document. The latter work builds on Sánchez et al. (2012), where sensitive terms are generalised.

14.5.7 Data Stream Anonymisation

A data stream is a sequence of data items that become available over time. This type of dynamic data is common in some environments, such as sensor networks, web logs, etc. Data streams are quite different from static data sets. In particular, streams are potentially infinite, may be fast flowing and may require fast processing for anonymisation. Because of these particularities, anonymisation methods that target dynamic data must be specifically designed. Whereas there is a large body of SDC methods for static data, the disclosure risk control literature on data streams is limited. The existing proposals follow three main approaches: perturbative masking, non-perturbative masking and counterfeiting.

In the perturbative masking approach, some noise is added to conceal the real value of the records. Li et al. (2007b) devised a method by which the correlation and the autocorrelation of multivariate data streams is tracked in an attempt to identify a good trade-off between privacy and utility. Differential privacy has also been used to anonymise data streams in some constrained scenarios. In Dwork et al. (2010), a differentially private counter of the number of 1's in a data stream is released at each step. This method was generalised in Bolot et al. (2013) to compute differentially private sums over restricted windows.

In the non-perturbative masking approach, one seeks to hide each record in the stream within a group of records. In the static data setting, k-anonymity and its extensions are well-known privacy models that follow this approach. In the work of

Cao et al. (2011a) and (2011b) these privacy models are adapted to streams. Since to make groups we need to accumulate records, this approach necessarily introduces some delay in the release of the anonymised stream. Quite recently, a perturbative adaptation of *k*-anonymity for streams, based on a primitive called steered microaggregation, has been introduced by Domingo-Ferrer and Soria-Comas (2017).

In the counterfeiting approach, a record is attempted to be hidden within a group of records. By hiding each record within a group of fake records, we avoid the delay inherent to the previous approach Kim et al. (2014). The main drawback is the overhead introduced by the addition of fake records.

14.5.8 Owner Privacy: Privacy-Preserving Data Mining

Privacy-Preserving Data Mining (PPDM) tries to solve the following question: *can we develop accurate data mining models without access to the data at the record level?* Therefore, it consists of techniques for modifying the original data in such a way that the private data remain private even after the mining process Verykios et al. (2004).

There are two radically different approaches to PPDM, namely, *PPDM based on perturbation* and *PPDM based on Secure Multiparty Computation* (SMC). The first was introduced by Agrawal and Srikant (2000) in the database community. Its idea is that respondents (who do not wish to reveal the exact value of their respective answers/records) or controllers (who wish to engage in joint computation with other controllers without disclosing their respective data sets to each other) compute modified values for sensitive attributes in such a way that accurate statistical results can still be obtained on the modified data. PPDM based on perturbation is largely based on statistical disclosure control techniques.

PPDM based on SMC, which was introduced by Lindell and Pinkas (2000) in the cryptographic community, addresses the problem of several entities holding confidential databases who wish to run a data mining algorithm on the union of their databases, without revealing unnecessary information. This type of PPDM is equivalent to data mining in distributed environments, where the data are partitioned across multiple parties. Partitioning can be vertical (each party holds all records on a different subset of attributes), horizontal (each party holds a subset of the records, but each record contains all attributes) or mixed.

Using SMC protocols based on cryptography (many of these resort to homomorphic encryption) or on sharing perturbed information in ways that do not alter the final results often requires changing or adapting the data mining algorithms. Hence, each cryptographic PPDM protocol is designed for a specific data mining computation and, in general, is not valid for other computations. For example, a secure scalar product protocol based on cryptographic primitives is applied to privacy preserving *k*-means clustering over a distributed dataset by Vaidya and Clifton (2003) and Jagannathan and Wright (2005). Similarly, Du et al. (2004) and Karr et al. (2009)

propose different ways (none of them based on encryption) to securely compute matrix products, which permits obtaining privacy-preserving linear regressions.

A different PPDM scenario arises when a data controller wants to leverage the storage and also the computational power of untrusted clouds to process her sensitive data. This setting was studied in the H2020 project 'CLARUS' (http://clarussecure.eu) and solutions based on cleartext data splitting across several clouds have been proposed. Furthermore, protocols to compute scalar products and matrix products with minimum controller involvement and maximum cloud involvement have been given by Domingo-Ferrer et al. (2018).

14.5.9 User Privacy: Private Information Retrieval

Finally, we address the privacy of the users querying a database. A history of queries to a database, or to a web search engine, can be used by the database owner to learn the interests of users, that is, to profile them. In this scenario, we seek to protect users from unrequested profiling by database owners. Mechanisms to achieve this goal are collectively referred to as private information retrieval (PIR).

Initial works on PIR, such as Chor et al. (1995), model databases as vectors of entries. Users requesting information from the database do so by providing an index or a set of indices of the database vector. In this setting, PIR techniques aim to hide the indices provided by the users. However, these initial approaches have several shortcomings. First, they require collaboration from the database owner, something that cannot be ensured unless database owners have a clear incentive to do so. Second, to perfectly hide the queried database indices one would need to query *all* entries in the database and then filter the results locally, which is clearly inefficient for moderately sized databases and certainly unfeasible for big databases. Finally, modelling a database as a vector and assuming that the user knows the indices where the desired information is stored is not applicable to most real databases, let alone web search engines.

Several solutions have been proposed to overcome such shortcomings. Domingo-Ferrer et al. (2009) propose a system named Goopir in which user queries are locally complemented with terms of similar frequency in the language (connected by OR operations). The responses are then filtered locally. TrackMeNot (Howe and Nissenbaum 2009) is a browser extension which periodically sends fake queries to web search engines so that the distribution of interests of the user is flattened and no useful profile can be extracted. Finally, other proposals such as the one by Reiter and Rubin (1998) make use of a P2P network in which users submit queries generated by other users to the web search engine, thus achieving the same results as TrackMeNot (flattened interest distributions) but without overloading the web search engines with fake queries.

14.6 Discrimination Prevention in Data Mining

Other than privacy implications, automated data collection and processing may have a secondary negative impact, which is discrimination. Automated data mining is used in several services to derive association and classification rules, which are then applied to a variety of decisions, such as loan granting, personnel selection, insurance premium computation, etc. While an automated classifier may be seen as a fair decision-making tool, if the training data are inherently biased, the generated rules will result in potentially discriminatory decisions.

Some works tackle this issue by pre-processing the training data using techniques akin to those from statistical disclosure control, but aimed at reducing the inherent bias in the data. Others act directly on the automatically mined rules, either by eliminating some of them or by generalising some of the conditions of these rules (Hajian and Domingo-Ferrer 2013; Hajian et al. 2014, 2015).

Acknowledgments and Disclaimer The following funding sources are gratefully acknowledged: European Commission (H2020–700540 CANVAS), Government of Catalonia (ICREA Acadèmia Prize to J. Domingo-Ferrer and 2017 SGR 705) and Spanish Government (project RTI2018–095094-B-C21 'Consent'). The views in this paper are the authors' own and do not necessarily reflect the views of UNESCO or any of the funders.

References

Abril D, Navarro-Arribas G, Torra V (2011) On the declassification of confidential documents. International conference on modeling decisions for Artificial Intelligence. Springer, pp 235–246

Agrawal R, Srikant R (2000) Privacy-preserving data mining. ACM 29(2). Available at: https://dl.acm.org/citation.cfm?id=335438. Last access 7 July 2019

Ben-Or M, Goldreich O, Goldwasser S, Håstad J, Kilian J, Micali S, Rogaway P (1988b) Everything provable is provable in zero-knowledge. In: Conference on the theory and application of cryptography. Springer, pp 37–56

Ben-Or M, Goldwasser S, Wigderson A (1988a) Completeness theorems for non-cryptographic fault-tolerant distributed computation. In: Proceedings of the twentieth annual ACM symposiumon Theory of computing. ACM, pp 1–10

Bier E, Chow R, Golle P et al (2009) The rules of redaction: identify, protect, review (and repeat). IEEE Secur Priv 7(6):46–53

Blanco-Justicia A, Domingo-Ferrer J (2018) Efficient privacy-preserving implicit authentication. Comput Comms (Elsevier) 125:13–23

Blum M, Feldman P, Silvio M (1988) Non-interactive zero-knowledge and its applications. In: Proceedings of the twentieth annual ACM symposium on Theory of Computing. ACM, pp 103–112

Bolot J, Fawaz N, Muthukrishnan S et al (2013) Private decayed predicate sums on streams. In: Proceedings of the 16th international conference on database theory. ACM, pp 284–295

Boneh D, Franklin M (2001) Identity-based encryption from the Weil pairing. Annual international cryptology conference. Springer, pp 213–229

Burridge J (2003) Information preserving statistical obfuscation. Stats Comput (Springer) 13(4):321–327

Cao J, Carminati B, Ferrari E et al (2011a) Castle: continuously anonymizing data streams. IEEE Trans Dep Secur Comput (IEEE) 8(3):337–352

Cao J, Karras P, Kalnis P et al (2011b) SABRE: a sensitive attribute Bucketization and REdistribution framework for t-closeness. VLDB J 20(1):59–81. Springer, New York

Chakaravarthy VT, Gupta H, Roy P et al (2008) Efficient techniques for document sanitization. In: Proceedings of the 17th ACM conference on information and knowledge management. ACM, pp 843–852

Chaum D (1983) Blind signatures for untraceable payments. Adv Cryptol:199–203

Chaum D (1981) Untraceable electronic mail, return addresses, and digital pseudonyms. Commun ACM 24:84–90

Chaum, Van Heyst E (1991) Group signatures. Workshop on the theory and application of cryptographic techniques. Springer, pp 257–265

Chaum D, Claude C, Damgaard I (1988) Multi-party unconditionally secure protocols. In: Proceedings of the twentieth annual ACM symposium on Theory of computing ACM, pp 11–19

Chor B, Goldreich O, Kushilevitz E, Sudan M (1995) Private information retrieval. Foundations of computer science, 1995. In: Proceedings., 36th annual symposium on IEEE, pp 41–50

Cumby CM, Ghani R (2011) A machine learning based system for semi-automatically redacting documents. IAAI. Available at: https://www.aaai.org/ocs/index.php/IAAI/IAAI-11/paper/view/3528. Last access 7 July 2019

Department of Health and Human Services, USA (1996) The Health insurance Portability and Accountability Act of 1996. Public Law:104–191

Dingledine R, Mathewson N, Syverson P (2004) Tor: The second-generation onion router. Naval Research Lab Washington DC

Domingo-Ferrer J, Solanas A, Castellà-Roca J (2009) H(k)-private information retrieval from privacy-uncooperative queryable databases. Online Inf Rev (Emerald Group Publishing Limited) 33(4):720–744

Domingo-Ferrer J, Soria-Comas J (2017) Steered microaggregation: a unified primitive for anonymization of data sets and data streams. Data Mining Workshops (ICDMW), 2017 IEEE International Conference on IEEE:995–1002

Domingo-Ferrer J, Mateo-Sanz JM (2002) Practical data-oriented microaggregation for statistical disclosure control. IEEE Trans Knowl Data Eng 14(1):189–201

Domingo-Ferrer J, Muralidhar K (2016) New directions in anonymization: permutation paradigm, verifiability by subjects and intruders, transparency to users. Inf Sci (Elsevier) 337:11–24

Domingo-Ferrer J, Torra V (2005) Ordinal, continuous and heterogeneous k-anonymity through microaggregation. Data Min Knowl Disc (Springer) 1(2):195–212

Domingo-Ferrer J, Wu Q, Blanco-Justicia A (2015) Flexible and robust privacy-preserving implicit authentication. IFIP International Information Security Conference. Springer:18–34

Domingo-Ferrer J, Sara Ricci S, Domingo-Enrich C (2018) Outsourcing scalar products and matrix products on privacy-protected unencrypted data stored in untrusted clouds. Inf Sci (Elsevier) 436:320–342

Domingo-Ferrer J, González-Nicolás U (2010) Hybrid microdata using microaggregation. Inf Sci (Elsevier) 180(15):2834–2844

Douglass MM, Cliffford GD, Reisner A et al (2005) De-identification algorithm for free-text nursing notes. Comput Cardiol (IEEE):331–334

Du W, Yunghsiang SH, Chen S (2004) Privacy-preserving multivariate statistical analysis: linear regression and classification. In: Proceedings of the 2004 SIAM international conference on data mining. SIAM, pp 222–233

Dwork C (2006) Differential privacy. Automata, Languages and Programming. 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II, pp 1–12

Dwork C, Naor M, Pitassi T, Rothblum GN (2010) Differential privacy under continual observation. Proceedings of the forty-second ACM symposium on Theory of computing. ACM:715–724

ElGamal T (1985) A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans Inf Theor (IEEE) 31:469–472

Finkel JR, Grenager T, Manning C (2005) Incorporating non-local information into information extraction systems by gibbs sampling. Proceedings of the 43rd annual meeting on association for computational linguistics. Assoc Comput Linguist:363–370

Gentry C (2009) Fully homomorphic encryption using ideal lattices. In: 41st ACM STOC, pp 169–178

Gentry C, Sahai A, Waters B (2013) Homomorphic encryption from learning with errors: conceptually-simpler, asymptotically-faster, attribute-based. In: Advances in cryptology – CRYPTO 2013. Springer, Berlin/Heidelberg, pp 75–92

Goldreich O, Micali S, Wigderson A (1987) How to play any mental game. In: Proceedings of the nineteenth annual ACM symposium on Theory of computing. ACM, pp 218–229

Goyal V, Pandey O, Sahai A et al (2006) Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on computer and communications security. ACM, pp 89–98

Groth J, Sahai A (2008) Efficient non-interactive proof systems for bilinear groups. Annual international conference on the theory and applications of cryptographic techniques. Springer, pp 415–432

Hajian S, Domingo-Ferrer J (2013) A methodology for direct and indirect discrimination prevention in data mining. IEEE Trans Knowl Data Eng 25(7):1445–1459

Hajian S, Domingo-Ferrer J, Farràs O (2014) Generalization-based privacy preservation and discrimination prevention in data publishing and mining. Data Min Knowl Disc (Springer) 28(5–6):1158–1188

Hajian S, Domingo-Ferrer J, Monreale D et al (2015) Discrimination- and privacy-aware patterns. Data Min Knowl Disc (Springer) 29(6):1733–1782

Hoepman J-H (2014) Privacy design strategies. IFIP international information security conference. Springer, pp 446–459

Howe DC, Nissenbaum H (2009) TrackMeNot: resisting surveillance in web search. Lessons from the identity trail: anonymity, privacy, and identity in a networked society, vol 23. Oxford University Press, pp 417–437

Hundepool A, Domingo-Ferrer J, Franconi L et al (2012) Statistical disclosure control. Wiley, Chichester

Jagannathan G, Wright RN (2005) Privacy-preserving distributed k-means clustering over arbitrarily partitioned data. In: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM, pp 593–599

Jakobsson M, Shi E, Golle P et al (2009) Implicit authentication for mobile devices. In: Proceedings of the 4th USENIX conference on Hot topics in security, pp 9–9

Jiang W, Murugesan M, Clifton C et al (2009) T-plausibility: semantic preserving text sanitization. Comput Sci Eng, 2009. CSE'09. International conference on. IEEE, pp 68–75

Karr AF, Lin X et al (2009) Privacy-preserving analysis of vertically partitioned data using secure matrix products. J Off Stat 25(1):125

Kim S, Sung MK, Chung YD (2014) A framework to preserve the privacy of electronic health data streams. J Biomed Inf (Elsevier) 50:95–106

Li F, Sun J, Papadimitriou S, et al (2007b) Hiding in the crowd: privacy preservation on evolving streams through correlation tracking. 2007 IEEE 23rd international conference on data engineering. IEEE, pp 686–695

Li N, Li T, Venkatasubramanian S (2007a) T-closeness: privacy beyond k-anonymity and ℓ-diversity. Data engineering, 2007. ICDE 2007. IEEE 23rd international conference on, pp 106–115

Lindell Y, Pinkas B (2000) Privacy preserving data mining. Annual International Cryptology Conference. Springer, pp 36–54

Machanavajjhala A, Gehrke J, Kifer D et al (2006) ℓ-diversity: privacy beyond k-anonymity. null:24

Muralidhar K, Sarathy R, Domingo-Ferrer J (2014) Reverse mapping to preserve the marginal distributions of attributes in masked microdata. International conference on privacy in statistical databases. Springer, pp 105–116

National Security Agency (2005) Redacting with confidence: how to safely publish sanitized reports converted from word to pdf. Available at: http://www.ca7.uscourts.gov/forms/nsa-redact.pdf. Last access 7 July 2019

Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. International conference on the theory and applications of cryptographic techniques, pp 223–238

Reiter MK, Rubin AD (1998) Crowds: anonymity for web transactions. ACM Trans Inf Syst Secur (TISSEC) (ACM) 1(1):66–92

Rubin DB (1993) Statistical disclosure limitation. J Off Stat 9(2):461–468

Safa NA, Safavi-Naini R, Shahandashti SF (2014) Privacy-preserving implicit authentication. IFIP international information security conference. Springer, pp 471–484

Samarati P, Sweeney L (1998) Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI International. Available at: http://www.csl.sri.com/papers/sritr-98-04/. Last access 7 July 2019

Sánchez D, Batet M, Viejo A (2012) Detecting sensitive information from textual documents: an information-theoretic approach. International conference on modeling decisions for Artificial Intelligence. Springer, pp 173–184

Sánchez D, Batet M, Viejo A (2013) Automatic general-purpose sanitization of textual documents. IEEE Trans Inf Forensic Secur 8(6):853–862

Shamir A (1984) Identity-based cryptosystems and signature schemes. Workshop on the theory and application of cryptographic techniques. Springer, pp 47–53

Shanqing G, Yingpei Z (2008) Attribute-based signature scheme. Information security and assurance, 2008. ISA 2008. International conference on. IEEE, pp 509–511

Sweeney L (1996) Replacing personally-identifying information in medical records, the scrub system. Proceedings of the AMIA annual fall symposium. Am Med Inform Assoc, p 333

Tveit A, Edsberg O, Rost TB et al (2004) Anonymization of general practitioner medical records. Second HelsIT Conference. Available at: https://pdfs.semanticscholar.org/c13b/fe9e6568c613f9e7a016a445bbc1372dd760.pdf. Last access 7 July 2019

Vaidya J, Clifton C (2003) Privacy-preserving K-means clustering over vertically partitioned data. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 206–215

Verykios VS, Bertino E, Fovino IN et al (2004) State-of-the-art in privacy preserving data mining. ACM Sigmod Rec 33(1):50–57

Yao AC-C (1986) How to generate and exchange secrets. Found Comp Sci, 1986., 27th annual symposium on. IEEE, pp 162–167