



HAL
open science

A Survey on Interdependent Privacy

Mathias Humbert, Benjamin Trubert, Kévin Huguenin

► **To cite this version:**

Mathias Humbert, Benjamin Trubert, Kévin Huguenin. A Survey on Interdependent Privacy. ACM Computing Surveys, 2020, 52 (6), pp.122:1-122:35. 10.1145/3360498 . hal-02044853

HAL Id: hal-02044853

<https://hal.science/hal-02044853>

Submitted on 3 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Survey on Interdependent Privacy

MATHIAS HUMBERT, Swiss Data Science Center, ETH Zurich and EPFL, Switzerland

BENJAMIN TRUBERT, University of Lausanne, Switzerland

KÉVIN HUGUENIN, University of Lausanne, Switzerland

The privacy of individuals does not only depend on their own actions and data but may also be affected by the privacy decisions and by the data shared by other individuals. This interdependence is an essential aspect of privacy and ignoring it can lead to serious privacy violations. In this survey, we summarize and analyze research on interdependent privacy risks and on the associated (cooperative and non-cooperative) solutions. We also demonstrate that interdependent privacy has been studied in isolation in different research communities. By doing so, we systematize knowledge on interdependent privacy research and provide insights on how this research should be conducted and which challenges it should address.

CCS Concepts: • **Security and privacy** → **Human and societal aspects of security and privacy**; **Economics of security and privacy**; **Social aspects of security and privacy**; **Privacy protections**;

Additional Key Words and Phrases: Interdependent privacy, Game theory, Statistical learning, User studies

ACM Reference Format:

Mathias Humbert, Benjamin Trubert, and Kévin Huguenin. 2019. A Survey on Interdependent Privacy. *ACM Comput. Surv.* 1, 1, Article 1 (February 2019), 39 pages. <https://doi.org/0000001.0000001>

1 INTRODUCTION

Privacy erosion in our connected world can be compared, in some ways, to climate change: Privacy degrades very gradually, which prevents global citizen awareness of the induced risks and rapid reaction from the people and societies being affected by this change. Back in 1992, in Rio, the international community met at the Earth Summit to take action on, among other dimensions, environment and climate change. In its preamble, the Rio Declaration recognizes the “*interdependent* nature of the Earth, our home.” Just like pollution and climate change affect countries in an interdependent manner, privacy issues affect individuals interdependently. In other words, there exist situations where the actions of some individuals affect the privacy of other individuals. Such situations are often referred to as interdependent privacy situations.

The simplest, yet striking, example of the interdependent nature of privacy is e-mails. As reported by Benjamin Mako Hill in a blog article, it is currently extremely challenging to keep full control and autonomy when it comes to e-mail privacy [Hill 2014]. Indeed, even though he ran his own e-mail server for 15 years, he realized that the majority of his e-mails were available to Google simply because his friends and contacts were using Gmail. Such a situation is at odds with privacy, defined as the right to control, edit, manage, and delete information about oneself and decide when,

Authors' addresses: Mathias Humbert, Swiss Data Science Center, ETH Zurich and EPFL, Lausanne, Switzerland, mathias.humbert@epfl.ch; Benjamin Trubert, University of Lausanne, Switzerland, benjamin.trubert@unil.ch; Kévin Huguenin, University of Lausanne, Switzerland, kevin.huguenin@unil.ch.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

Manuscript submitted to ACM

how, and to what extent information is communicated with others [Westin 1970]. The recent Facebook-Cambridge Analytica scandal further demonstrates that our own privacy behavior can affect our friends' privacy [Cadwalladr and Graham-Harrison 2018]. Indeed, Cambridge Analytica exploited Facebook (FB)'s application settings not only to gather information of users who accepted to take their survey (and be paid for it), but also to gather the personal information of all their friends who had never granted access to their own data.

These examples demonstrate the extent of the concrete threat on people's personal data and privacy in today's highly connected world, and they shed light on a serious risk that has so far been marginally studied, in a fragmented way, by the scientific community. In general, such risks stem from the fact that data features multiple individuals or the fact that data that seems to involve only certain individuals in fact reveals information about others. Hence, these risks call for specific solutions. In this article, we survey the literature on the situations where the actions of some individuals affect the privacy of others, and we systematically analyze the works on the associated risks and solutions. We also analyze how research in this area has been conducted from different communities. More precisely, the contributions of our survey are the following:

- We review the various notions, definitions, and terminologies that are related to privacy and interdependent privacy and that have been used in different research communities. We further provide a graph-based analysis of the literature; we highlight that research on the topic has been conducted in isolation in different communities (Sec. 2).
- To enable the reader to understand existing works and to produce new ones, we provide some technical background, in the form of a tutorial, on the standard theories and tools used in interdependent privacy research (Sec. 3).
- We survey the existing literature on interdependent privacy *risks and concerns*, and we categorize them in a systematic way based on (among other things) the data types considered and the techniques used. (Sec. 4).
- We survey the existing literature on (technical and non-technical) solutions for preventing or mitigating interdependent privacy risks, and we categorize them in a systematic way (Sec. 5).
- We issue recommendations on how interdependent privacy research could be conducted and on the issues that should be addressed (Sec. 6).

2 DEFINITIONS

In this section, we introduce the notions and definitions relevant to the topic of interdependent privacy, and we report on our analysis of how interdependent privacy has been referred to and studied in different research communities.

2.1 Privacy

Westin defines privacy as the right to control, edit, manage, and delete information about oneself and to decide when, how, and to what extent information is communicated to others [Westin 1970]. Interestingly, in this definition, the word "information" is used (and not data); this means that this definition applies to any data that reveals information about the considered individual. According to this definition, every individual should be able to manage and protect information about themselves, individually and independently from the others. This is unfortunately not always possible, either because an individual does not have the full control over the sharing policies (i.e., to which audience the data is revealed) of the data or because of the intrinsic properties of the data (i.e., correlation between the data of different individuals or data revealing information about multiple individuals).

Nissenbaum defines privacy as contextual integrity, where each context is associated with social norms [Nissenbaum 2010]. In her definition, the five critical parameters that define contextual informational norms are the data subject, the sender of the data, the recipient of the data, the information type, and the transmission principle. As shown in this

survey, among these parameters, the data subject, the sender of the data, and the recipient of the data could be subject to interdependent privacy issues, which raises new challenges to be addressed.

The case where the actions of some individuals affect the privacy of others has been increasingly studied over the last decade, by different communities using different terminologies – in isolation as we will see. Hereafter we list the definitions used in the literature of these communities. We summarize the results in Table 1. In this article, including its title, we use the term interdependent privacy by analogy with the general concept of interdependent security and because we believe it describes best the considered situations.

2.1.1 Collective Privacy. Squicciarini et al. were the first to study the problem of collaborative management of privacy settings for content that involves multiple users of an online social network (OSN) [Squicciarini et al. 2009]. They refer to this problem as *collective privacy* and give photo sharing on OSNs as an example.

2.1.2 Multi-Party Privacy. Thomas et al. tackle the problem of conflicting privacy settings between friends on OSNs [Thomas et al. 2010]. They refer to this problem, and more generally to the situations where more than one OSN user controls the visibility of some shared content, as *multi-party privacy*. Later, Hu and Anh also used the term *multiparty* authorization framework to refer to a new access control model for collaboratively managing shared content in OSNs [Hu and Ahn 2011].

2.1.3 Networked Privacy. Boyd coined the term *networked privacy* to refer to data and privacy practices that affect not only the individual who shares the data but also other individuals, i.e., those related to the data. For example, she mentions genomic data and photos shared online without the consent of the involved individuals [Boyd 2012].

2.1.4 Interdependent Privacy. Biczók and Chia study situations where the privacy of users of OSNs is affected by the actions of others – essentially granting access to their friends’ data to third-party applications [Biczók and Chia 2013]. They refer to such situations as *interdependent privacy* situations. This terminology was later used by Humbert et al., Pu and Grossklags, and Olteanu et al.. Extending this terminology, Olteanu et al. coined the term *interdependent personal data* to refer to the case where data (seemingly) concerns a single individual but is, in fact, related to others because of data correlation [Olteanu et al. 2018].

2.1.5 Peer Privacy. Chen et al. coined the term *peer disclosure* to refer to the disclosure of an individual’s private information by peers on OSNs [Chen et al. 2015]. In their experiments, they use photo tagging as an example of peer disclosure. Later, Ozdemir et al. use the terms *peer privacy* and *privacy concerns in a peer context* to refer to the same situations [Ozdemir et al. 2017].

2.1.6 Group Privacy. Bloustein initially coined the term *group privacy* to refer to the fact that the actions of individuals surrounding an individual (friends, family, strangers in the public space, etc.) can affect this individual’s privacy [Bloustein 1978]. This term is used by Radaelli et al., with the same meaning [Radaelli et al. 2018]. However, it is also used in the literature to refer to the right to keep an individual’s affiliation to a group private [Floridi 2014; Narayanan and Shmatikov 2005]. Group privacy is also used in the differential privacy terminology when protecting databases that differ in more than one element [Dwork et al. 2006].

2.1.7 Multiple-Subject Privacy. Gnesi et al. use the term *multiple-subject personal data* to define personal data related to multiple subjects where each subject should hold rights to control how the data is shared [Gnesi et al. 2014]. They give records of phone calls, co-locations, and reports of medical examinations as examples.

Table 1. Terminology used to refer to interdependent privacy.

terminology	original article	other articles
collective privacy	[Squicciarini et al. 2009]	[Ilija et al. 2017; Jia and Xu 2016a,b; Ratikan and Shikida 2014; Squicciarini et al. 2010, 2011]
multi-party privacy	[Thomas et al. 2010]	[Fogues et al. 2015; Hu and Ahn 2011; Hu et al. 2013; Li et al. 2017a; Such and Criado 2014, 2016, 2018; Such et al. 2017; Such and Rovatsos 2016]
networked privacy	[Boyd 2012]	[Cho and Filippova 2016; Marwick and Boyd 2014; Vitak et al. 2015]
interdependent privacy	[Biczók and Chia 2013]	[Ayday and Humbert 2017; Harkous and Aberer 2017; Humbert et al. 2017; Olteanu et al. 2018, 2017, 2019; Pu and Grossklags 2014, 2015, 2016, 2017; Symeonidis et al. 2016a; Weidman et al. 2018]
peer privacy	[Chen et al. 2015]	[Ozdemir et al. 2017]
group privacy	[Bloustein 1978]	[Radaelli et al. 2018]
multiple-subject privacy	[Gnesi et al. 2014]	[Olteanu et al. 2018]

2.2 Analysis of Research Communities

In order to understand the origin and the use of the different terminologies that refer to interdependent privacy situations, we conduct an analysis of the relevant literature.

Methodology. We first build the citation graph of the articles (referenced in this survey) that deal directly with interdependent privacy. A vertex of the graph represents an article, and there is an edge from an article to another if the former references the latter. For each article, we use its digital object identifiers (DOI), title, and authors' names to obtain its list of references from the Microsoft Academic database.¹ In order to identify, from the structure of the citation graph, the different communities that tackle the problem of interdependent privacy, we rely on a clustering algorithm executed on the undirected version of the citation graph (i.e., there is an edge between two articles if one of them cites the other one). More specifically, we use the Louvain method for community detection [Blondel et al. 2008]: It does not require us to specify the number of clusters a priori (it is determined automatically by the algorithm), it takes as input an undirected graph and it relies only on the structure of the graph (i.e., it is oblivious of the semantic of the vertices). We also identify the most central articles according to the betweenness centrality metric that is related to the number of shortest paths that go through a particular vertex.

Results. The source code of the Python program used to perform the analysis and the data (i.e., the citation graph) are available online.² The graph contains a total of 93 articles and 325 edges (≈ 3.5 citations per article). The community-detection algorithm finds four communities plus eleven isolated articles. These communities correspond to siloed research tracks in interdependent privacy from the following research domains: information security and privacy (ISP, 19 articles), human-computer interaction/computer-supported cooperative work (HCI/CSCW, 19 articles), data science (DS, 22 articles), information system (IS, 22 articles). In the ISP community, the majority of the articles were published after (and cite) the article of Biczók and Chia, which coined the term “interdependent privacy”. Although these communities are somewhat isolated, some articles are connecting them. In general, the articles with a high centrality are either articles that connect communities or articles that start a research track (and coined the associated terms).

¹<https://academic.microsoft.com>, last visited: Feb. 2019

²https://github.com/isplab-unil/survey_interdependent_privacy

3 BACKGROUND

In this section, we give some background about the standard theories and techniques used to study interdependent privacy situations. Note that these theories and techniques were not necessarily introduced in the context of interdependent privacy. Readers familiar with some or all of the theories and techniques can skip the associated subsections. These theories and techniques were identified from our analysis of the literature on interdependent privacy. They can be used either in an offensive way by an adversary (e.g., statistical inference) or by individuals in a defensive way (e.g., cryptography and access control), or for analysis purposes (e.g., game theory and communication privacy management). For each theory and technique, we explain their main concepts, we provide, when applicable, pointers to the software tools that implement them, we illustrate them on sample use cases extracted from the interdependent privacy literature, and we explain why and how they help study interdependent privacy situations.

3.1 Statistical Inference

We present an important type of statistical inference methods, Bayesian inference: It is extensively used in interdependent privacy scenarios to update the knowledge (probability) about some attribute(s), typically of the target individual, given the observation of other attributes, typically those of other individuals (see examples below). With Bayesian inference, probabilities are interpreted as the quantification of a belief/knowledge; it relies on the Bayes' theorem that links conditional and marginal probabilities:

$$P(\text{attributes} \mid \text{observations}) = \frac{P(\text{observations} \mid \text{attributes}) \cdot P(\text{attributes})}{P(\text{observations})}, \quad (1)$$

where the marginal probability $P(\text{attributes})$ is called the prior and the conditional probability $P(\text{attributes} \mid \text{observations})$ is called the posterior (i.e., the prior updated after making the observations).

In order to efficiently perform Bayesian inference, researchers have developed probabilistic graphical models (PGMs) [Koller and Friedman 2009], graph-based models that enable the clear (graphical) representation of conditional or joint (in)dependencies between various attributes (represented by random variables). Bayesian networks are used to represent directed dependencies, i.e., conditional probability distributions between random variables, whereas Markov random fields are used for undirected dependencies, i.e., joint probability distributions between random variables. Hidden Markov models (HMMs) are a special instance of Bayesian networks.

Belief propagation [Pearl 1988] is a well-known message-passing algorithm for performing Bayesian inference on PGMs. It has been extensively used in artificial intelligence and information theory, e.g., in low-density parity-check codes. The belief-propagation algorithm can be applied to efficiently compute marginal distributions of unknown variables given a set of observed variables connected to them in the underlying PGM. Belief propagation and PGMs jointly allow to reduce the complexity of marginalization from exponential in the number of random variables down to linear (depending on the underlying PGM structure).

Figure 1 depicts two concrete models of Bayesian networks that were used in interdependent privacy settings for performing inference on (a) genomic data and (b) location data. Software libraries implementing Bayesian networks operations include bnt (Matlab),³ Netica (Java, C, C++, etc.),⁴ and pgmpy (Python),⁵ to name a few.

³<https://github.com/bayesnet/bnt>, last visited: Feb. 2019.

⁴<https://www.norsys.com/>, last visited: Feb. 2019.

⁵<https://github.com/pgmpy/pgmpy>, last visited: Feb. 2019.

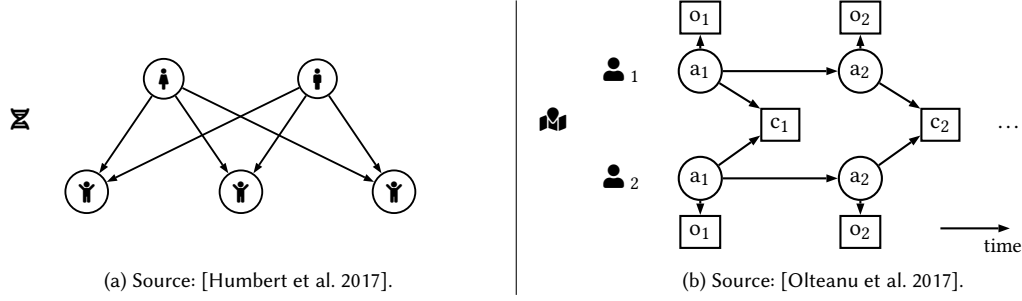


Fig. 1. (a) Bayesian network used for genome inference in a family tree with 2 parents (top) and 3 children (bottom). The genome of an individual depends only on that of their parents. (b) Bayesian network used for location inference. a_t represents the actual location of an individual (at time t), o_t represents the obfuscated location observed by the adversary and c_t the co-location of two individuals; the obfuscated location depends only on the actual location; the co-location depends only on the actual location of the involved individuals. Under the Markov assumption, the actual location at time t depends only on the location at the previous time instant.

Application to interdependent privacy. The privacy of an individual is often quantified as the error an adversary makes when inferring the values of some of the individual’s private attributes. Therefore, statistical inference techniques are widely used in privacy risk assessment. In particular, in interdependent privacy scenarios where statistical dependencies exist between the target individual’s data and that of their contacts (e.g., relatives, friends, co-workers), PGMs are very relevant. Many works on interdependent privacy risks assessment rely on Bayesian inference and/or Bayesian networks or Markov random fields (e.g., [Backes et al. 2018; Berrang et al. 2018; He et al. 2006; Humbert et al. 2017; Jia et al. 2017; Olteanu et al. 2017; Sadilek et al. 2012]) (see Sec. 4 and more specifically Table 2 on page 22).

3.2 Game Theory

Game theory is the study of the interaction between multiple rational decision makers (referred to as *players*) who aim to optimize their utility (referred to as *payoff*) [Myerson 2004; Von Neumann and Morgenstern 2007]. It has been successfully applied to fields such as economics, biology, and political science, but also in computer science and more specifically computer security and privacy [Laszka et al. 2014; Manshaei et al. 2013]. A key concept of game theory is the notion of *equilibrium*, more precisely Nash equilibrium (NE), that enables the modeling and prediction of stable states where, given other players’ strategies, no player has an incentive to deviate from his strategy. The optimal strategy given others’ strategies is formalized with the so-called *best response*. Mathematically, the best response $br_i(s_{-i})$ of player i to the profile of others’ strategies s_{-i} is defined as follows:

$$br_i(s_{-i}) = \arg \max_{s_i \in \mathcal{S}_i} u_i(s_i, s_{-i}), \quad (2)$$

where $u_i(s_i, s_{-i})$ represents player i ’s payoff, given the strategy profile $s = (s_i, s_{-i})$ encompassing all players.⁶ From this definition, the notion of NE can be defined as the case where all players’ strategies are the mutual best responses to each other. Formally, the strategy profile s^* is a NE if, for each player i ,

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \forall s_i \in \mathcal{S}_i \quad (3)$$

There exist two main types of games: simultaneous games (or static games) and sequential games (or dynamic games). In simultaneous games, players move at the same time and are not aware of the move(s) of the other(s), whereas in

⁶In game theory, $-i$ is a common notation to represent all players but i .

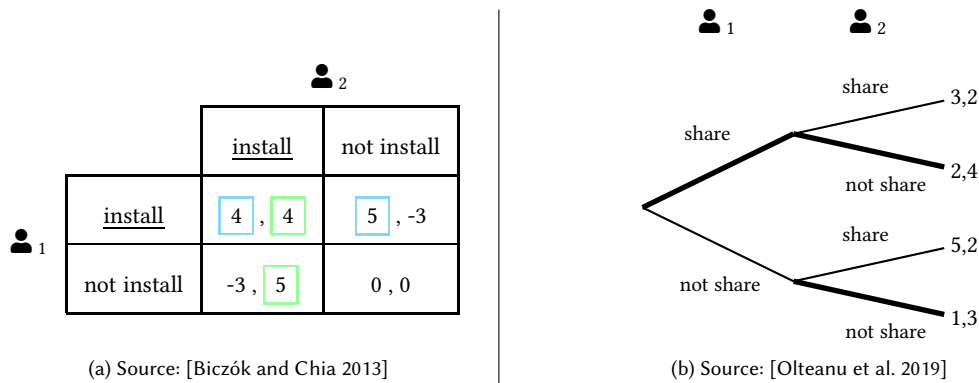


Fig. 2. Two-player information sharing game where each player chooses between two strategies (install app vs. not install app or share information vs. not share information): (a) a simultaneous game represented in its normal form. Both players play at the same time. Each cell of the matrix represents the respective payoffs of the players for the corresponding strategies; the best responses are framed and the dominant strategies are underlined. (b) a sequential game represented in its extensive form. The second player plays after the first player. The values at the leaves (right) represent the respective payoffs of the players for the corresponding sequence of strategies. The thick lines represent the best moves in the backward induction procedure.

sequential games players have some knowledge about the other players' previous moves. A main difference between these game types is in their representation (see Figure 2). Normal-form (also called strategic-form) representation is used for simultaneous games (see Figure 2a), whereas extensive-form representation is used for sequential games (see Figure 2b). The normal form is represented by a matrix of payoff of all players for all strategy profiles, whereas the extensive form is represented by a decision tree. Finally, repeated games are an extension of sequential games and, as such, also use the extensive-form representation. They consist in the repetition of a single-stage game multiple times and capture the effect of players on the future payoffs and moves of other players (which is a key feature of the extensive form, as discussed below).

It should be noted that the extensive form contains more information than the normal form, e.g., the sequence of players. Therefore, there can be multiple extensive-form games corresponding to the same normal form. Besides, more subtle notions of NE are needed for reasoning about extensive-form/sequential games. For instance, a sub-game perfect Nash equilibrium (SPNE) is an equilibrium derived by considering a smaller part of the whole game tree and by eliminating *incredible risks* (i.e., strategies that would not rationally be chosen). A common method for finding SPNE in finite games is referred to as *backward induction*. It first considers the last actions of the game and derives the best decision of the last player given all other previous possible decisions in the game. It then considers the second to last actions and again derives the utility-maximizing decision of the corresponding player. The same process is applied to all tree levels, until one reaches the root of the tree.

Figure 2 depicts simple examples of the two main game types discussed above. Figure 2a shows a two-player simultaneous game with (the same) two strategies for each player. We can immediately notice that the “install” strategy is the best response for both players, regardless of the other player’s strategy. Strategy “not install” is said to be *strictly dominated* by “install”. In this example, (“install”, “install”) is the unique NE. Figure 2b shows a two-player sequential game with (the same) two strategies for each player. Using backward induction, we first define the second player’s best moves given the possible history. If the first player (Player 1) chooses “share”, Player 2’s best move is to not share, as 4 is larger than 2. Similarly, if Player 1 decides to not share, then Player 2’s best move is to not share either, as 3 is larger

than 2. Then, given the best moves of Player 2, player 1 chooses her best moves too, which is to share ($2 > 1$). Hence, they reach a NE (“share”, “not share”).

Software libraries implementing game-theoretical analysis tools (e.g., backward induction and finding NE) include Gambit (Python binding)⁷ and MatTuGames (MATLAB),⁸ to name a couple.

Application to interdependent privacy. Because game theory can well model and capture the interactions between individuals with potentially conflicting interests, it is very helpful for predicting the strategies adopted by individuals and the stable states in interdependent privacy settings. Various game-theoretic models have been used in the context of interdependent privacy, such as simultaneous games [Biczók and Chia 2013; Humbert et al. 2015], sequential games [Olteanu et al. 2019] and repeated games [Hu et al. 2014] (see Sec. 5.1.1 and more specifically Table 3 on page 24).

3.3 Cryptography

Cryptography refers to the process of transforming intelligible information (plaintext) into unintelligible one (ciphertext) [Menezes et al. 1997]. Modern cryptography can be split into two main fields: symmetric and asymmetric cryptography. Symmetric-key cryptography (e.g., DES, AES) refers to the case where both the sender and the receiver of some information share the same key, while asymmetric cryptography (e.g., RSA, ECC) relies on two different keys, a public key that can be accessed by anyone, and a private key, that should be known only to the receiver (to ensure confidentiality of the message) or the sender (to ensure message integrity or sender authentication).

Secret sharing is a cryptographic scheme that enables a group of users to protect a secret. It can be used to protect a symmetric key among n users for instance. Each user receives a share in such a way that any subset of t or more users can reconstruct the secret, but no information can be learned from less than t shares. One of the most popular secret sharing scheme was proposed by Shamir; it is based on interpolation of a polynomial of degree $t - 1$ [Shamir 1979].

Attribute-based encryption is a public-key crypto-system that allows for a fine-grained sharing of encrypted data [Goyal et al. 2006; Sahai and Waters 2005]. In such a crypto-system, the ciphertext and/or the user’s secret key are labeled with a set of descriptive attributes, and decryption of the ciphertext is possible only by users who hold a set of attributes that match the rules included upon encryption.

Homomorphic encryption is a class of cryptographic methods that allows computations directly on the ciphertexts and for which only the results of the computations can be returned in the plaintext domain. This approach can be especially useful in the context where a central entity (e.g., service operator) is not trusted. Notable examples of (partially) homomorphic encryption schemes are the Paillier [Paillier 1999] and ElGamal [ElGamal 1985] crypto-systems.

Application to interdependent privacy. In general, cryptography can be helpful in the case of interdependent privacy (mostly for co-owned data as cryptographic techniques are data agnostic, and therefore cannot address data correlation issues per se), to hide information to certain adversarial parties. For instance to share content online or decide on the visibility of some shared content without the service providers and unauthorized users having access to it [Beato and Peeters 2014; Ilia et al. 2015; Olteanu et al. 2018; Palomar et al. 2016] (see Sec. 5.2 and more specifically Table 4 on page 31). More specifically, in the case of co-owned data and multi-party privacy conflicts, cryptography can be used to ensure that some content can only be accessed with the consent of the co-owners. For this purpose, cryptographic secret-sharing techniques (used by Ilia et al. [Ilia et al. 2015]) are particularly well suited.

⁷<http://www.gambit-project.org>, last visited: Feb. 2019.

⁸<https://ch.mathworks.com/matlabcentral/fileexchange/35933-mattugames>, last visited: Feb. 2019.

3.4 Communication Privacy Management Theory

Communication privacy management (CPM) theory, developed by Petronio, models how individuals share and manage private information when communicating with each other [Petronio 2002]. This theory has been successfully applied to analyze information sharing in all sorts of relationships, including between relatives (e.g., parent-child), friends, and co-workers (e.g., superior-subordinate), be it offline (i.e., in the real world) or online (i.e., on the Internet). This theory builds on the metaphor of boundaries that separate individuals' private (i.e., private boundary) and public spheres. CPM theory is based on five key principles: ownership, control, rules, co-ownership and boundary turbulence [Petronio 2010]. Private information belongs to a particular individual defined as the owner. When private information is shared, a collective boundary is created between the individual who discloses the information and the one who receives it, the receiver (e.g., a tagged individual). A receiver is then considered as a co-owner of the shared information and a collective boundary protects the private information by defining privacy rules to coordinate all the private boundaries of co-owners. These rules control who has access to the information and determine the possibility of extending the involved co-owners (linkage rules), the degree of access to the information (permeability rules), and the implication of co-owners in the decision making of rules (ownership rules). The last of the five principles, boundary turbulences, occurs when a co-owner discloses, by mistake or intentionally, the information outside the initially defined boundary.

Application to interdependent privacy. Communication privacy management theory helps model how individuals collaboratively decide on the audience of some content they co-own on OSNs, a typical interdependent privacy situation / multi-party privacy conflict. In such cases, permeability rules capture fine-grained access control (e.g., show only part of the faces in a picture), and boundary turbulences capture the increase of the audience of a content when a user re-shares some content, thus making it visible and accessible to their own contacts on the OSN. It has been applied in several works [Chen et al. 2015; Choi and Jiang 2013; Jia and Xu 2016b] (see Sec. 5.1.2).

3.5 Access Control

Access control (AC) techniques control the access to some resources by some entities (e.g., a user) via a mechanism (e.g., a reference monitor) that decides, based on a set of rules also called policies, to grant or deny access to the resources. In the context of privacy protection, the resource to which access is controlled is typically sensitive data. Users can be authenticated by: something they know (e.g., a password or a key), something they have (e.g., a smart-card), and something they are (e.g., fingerprint). Several models exist to define access based on the identity or on the role of users, such as mandatory access control (MAC), discretionary access control (DAC), and role-based access control (RBAC). Relationship-based access control (ReBAC) improves upon the RBAC model by adding granularity and context in the definition of roles [Fong 2011]. ReBAC relies on social network data to determine the type of relationship (if any) between two users. The most common AC policies when sharing new content on OSNs are: select specific users, select specific groups of users/friends, or select pre-defined categories of users (e.g., friends of friends and public). In the cases where the data can be divided into sub-resources (e.g., faces in a group picture), access control rules are determined for each sub-resource. An important challenge, in collaborative systems, is to combine the access control rules of the shared resource based on the individual rules and preferences of the involved users.⁹ Typical combination strategies include voting (e.g., majority voting and veto voting).

⁹Paci et al. wrote a comprehensive survey on collective access control solutions [Paci et al. 2018]; note, however, that this survey does not cover specifically (interdependent) privacy context but collaborative systems in general.

Application to interdependent privacy. Access-control techniques are particularly useful in interdependent privacy situations as they help determine the audience of co-owned data, typically shared online. Such techniques have been used in several works [Hu et al. 2011; Rathore and Tripathy 2016; Wishart et al. 2010; Xu et al. 2017]. ReBAC has also been successfully used in the case of OSNs where social relationship information is available (e.g., [Mehregan and Fong 2016]) (see Sec. 5.2 and more specifically Table 4 on page 31).

4 PRIVACY RISKS AND CONCERNS

In this section, we survey the concrete interdependent privacy risks (i.e., situations where the actions of some individuals affect the privacy of others) studied in the literature. In most of the considered risks, “actions” refer to information disclosure, typically on online platforms. The root cause of such risks is either that the data disclosed directly features multiple individuals or that the private attributes of different individuals, hence the data associated with them, are often times correlated, especially when these individuals are somehow related (e.g., friends, family members).¹⁰ Two typical sources of correlation are (1) homophily for friends (the fact that individuals tend to befriend with individuals with whom they share similarities) and (2) genetic inheritance (the fact that individuals inherit their genetic material from both their parents). The source and the nature of correlation highly depend on the type of data considered.

Correlation often indicates a predictive relationship between the private attributes of individuals; these relationships can be exploited, by an adversary, at the expense of these individuals’ privacy. Indeed, by considering the statistical properties of the relationship between the data of different individuals, the unknown private attributes of an individual can be inferred, or, more generally, the knowledge about them can be refined, algorithmically from the data disclosed by other individuals. Beyond correlation, interdependent privacy risks can also arise when individuals disclose, intentionally or not, to third parties the information related to another individual, which they were trusted with in the first place.

Privacy leakage is related to the increase in the knowledge acquired by the adversary from the disclosed data, possibly through inference. In interdependent privacy situations, the leakage comes also from other individuals. It is important to note that interdependent privacy refers to a *class* of situations, not to a specific privacy metric; in fact, standard metrics for quantifying/defining privacy are used to quantify interdependent privacy risks. These include k -anonymity [Sweeney 2002] (and its refinements, e.g., l -diversity [Machanavajjhala et al. 2006]) w.r.t. certain pseudo identifiers and private attributes of the target individuals, differential privacy [Dwork et al. 2006] w.r.t. certain private attributes and queries, and uncertainty (usually captured by Shannon’s entropy) and incorrectness (usually captured by the distance between the ground truth and the inferred value, i.e., the expected error) w.r.t. certain private attributes. In other words, because of the actions/data of other individuals, an individual’s privacy might be decreased, which would result in a decrease of k (for k -anonymity), an increase of ϵ (for ϵ -differential privacy),¹¹ or a decrease in entropy/expected error. Such metrics, however, are not specific to interdependent privacy; Wagner and Eckhoff wrote a comprehensive survey on privacy metrics including for social network, genomic and location data [Wagner and Eckhoff 2018]. They conclude from their survey that multiple metrics are often needed to cover multiple aspects of privacy, and they propose a series of nine questions to guide the selection of appropriate metrics for a given scenario. Furthermore, in the supplementary materials, they briefly cover interdependent privacy and mention that existing privacy metrics can be relied upon to quantify the privacy effect of interdependence. Most works presented in this section rely on uncertainty and incorrectness for quantifying privacy.

¹⁰Mondal et al. discuss and formalize the differences between accessibility, visibility and inferrability of private information on OSN in the context of interdependent privacy situations [Mondal et al. 2014].

¹¹In fact, data correlation can even void the privacy guarantees provided by differential privacy as it violates one its core underlying assumption: data independence. This was first shown by Kifer and Machanavajjhala [Kifer and Machanavajjhala 2011]; we explain this in Sec. 4.5

We structure this section according to the types of data involved in the different risks surveyed. For each data type, we first describe the context and give the necessary background, before we present the different risks associated with the considered data type. A typical risk consists in inferring (or refining knowledge on) certain private attributes of an individual, based on the data disclosed by other individuals, assuming that the adversary has information about the relationship between individuals.¹² When there are experiments, we also present the experimental setup and results, as well as the associated findings. We also present, for some data types, studies on the associated user privacy concerns.

We summarize the results of this section in a synthetic table presented at the end of this section (Table 2, on page 22).

4.1 Demographics and Preferences Data

Context. Studies on real-life social networks have demonstrated the tendency of individuals to bond with individuals who share certain characteristics with them. Such characteristics include gender, age, social class, geographic location, and preferences. This phenomenon is known as *homophily* (or through the saying, “birds of a feather flock together”), a term coined by McPherson et al. [McPherson et al. 2001] and is well documented in the sociology community. Naturally, homophily introduces a correlation between the personal data of individuals bound by social ties. This correlation can be exploited for inference purposes, thus creating interdependent privacy risks.

Such risks have materialized on a large scale with the proliferation of OSNs (e.g., FB and Twitter) where individuals disclose their demographic attributes and preferences (characteristics known to be covered by homophily), as well as social ties (e.g., friends, colleagues). The disclosure of such social network data – known as user profiles – is often times made in a structured way, through specific data fields for which the values must be selected in predefined lists, thus facilitating the algorithmic processing of the data. Furthermore, social network data can be made available to various entities beyond the OSN operators themselves, including other users of the OSNs, third parties (e.g., developers) or even the public. Beyond the aforementioned privacy risks based on homophily, the simple fact that users reveal information to their friends (e.g., on OSNs) creates new privacy risks because their friends can give away this information to other entities, including other users (e.g., their own friends) and third parties (e.g., developers of apps for OSNs). In fact, Facebook’s data-use policy contains a paragraph named “I. What kinds of information do we collect?→Things you and others do and provide→Things others do and information they provide about you”¹³ which states:

“We also receive and analyze content, communications and information that other people provide when they use our Products. This can include information about you, such as when others share or comment on a photo of you, send a message to you, or upload, sync or import your contact information.”

And Facebook aggregates the information provided by multiple users about some individual in the form of a so-called shadow profile (even for non-Facebook users).¹⁴ OSN data has been proven to be used as a basis for discrimination by (potential) employers (e.g., [Acquisti and Fong 2012]), identity theft, and cyberstalking, hence it is quite sensitive.

Risks. He et al. were the first to identify and formalize the interdependent privacy risks related to OSN data where personal information about a user¹⁵ can be inferred from the data of their contacts (often referred to as “friends”) [He

¹²Several works demonstrate how relationships (e.g., friends, family members, colleagues) – or in general correlation between individual data – can be inferred from various types of data including location [Backes et al. 2017; Bilogrevic et al. 2013; Crandall et al. 2010; Eagle et al. 2009; Sadilek et al. 2012], genomic data [Berrang et al. 2018; Kale et al. 2018; Manichaikul et al. 2010; Thornton et al. 2012] and photos [Shoshitaishvili et al. 2015]. In fact, the inference works both ways: Personal data can be used to infer relationships and relationship data can be used to infer (or refine knowledge on) personal attributes (e.g., [Gong et al. 2014; Sadilek et al. 2012]). Works on relationship inference, however, are outside of the scope of this survey.

¹³https://www.facebook.com/full_data_use_policy, last visited: Jun. 2019.

¹⁴This practice was extensively discussed from the early 2010’s to recently, especially in the context of the data leak that affected the “Download Your Information” program [noa 2013; Blue 2013a,b; Quodling 2018].

¹⁵Note that the target individuals might not even be users of the considered OSNs. But they might be identifiable, e.g., based on their names.

et al. 2006]. They formalize the problem and provide the technical foundations, based on Bayesian networks (see Sec. 3.1), for assessing privacy risks in such situations. In particular, they consider both single-hop and multi-hop inference: the cases where (1) the information of only the target’s friends is used and (2) where the information about the friends of the target’s friends (and so-on) is used. They evaluate the performance of the proposed inference techniques – hence the magnitude of the associated privacy risks – by relying on a dataset of $N \approx 67k$ users of the LiveJournal online weblog with real ties and synthetic attribute values. Their results show a substantial increase in accuracy (from 70% to up to 95%) when information about (friends of) friends is used.

Becker and Chen focus on generic attribute inference based on the attributes of users’ friends [Becker and Chen 2009]. More specifically, users are assigned a given value for one of their attributes if at least a certain number of their friends have this value for the considered attribute. The authors focus their work on FB and consider 12 different attributes including age, relationship status, and zipcode. They propose a tool named PrivAware to enable users to assess and to reduce the accuracy of such an inference.

Lindamood et al. focus on (binary) political orientation (conservative vs. liberals) inference [Lindamood et al. 2009]. They evaluate the performance of an ad-hoc inference algorithm based on a naïve Bayes classifier: It uses the data publicly disclosed by the individuals who are bound with the target individuals through a social tie (called “friends”) declared on the OSN. To do so, they rely on a sample of $N \approx 35k$ FB profiles from the Dallas/Forth Worth network.

Zheleva and Getoor study the same problem but also study gender inference and consider more fine-grained political orientations (i.e., six labels instead of two) [Zheleva and Getoor 2009]. In addition, they design an algorithm that further exploits the notion of social groups in the inference process. They evaluate the performance of the algorithms on four different datasets including a sample of $N \approx 1.5k$ FB profiles in the US for political orientation ($N \approx 1k$ for gender).

Jernigan and Mistree focus on sexual orientation inference [Jernigan and Mistree 2009]. By analyzing a dataset of $N \approx 6k$ FB user profiles of university students, they found that the proportion of gay male friends was significantly higher for gay males than for other individuals. By relying on a logistic regression-based classifier, they demonstrate that this can be used to accurately infer a male individual’s sexual orientation.

Sarigol et al. also study sexual orientation inference for both users and non-users¹⁶ of OSNs [Sarigol et al. 2014]. They focused on a dataset of Friendster profiles ($N \approx 1M$) in the US. Their results show a high accuracy and that the privacy leakage depends on the size of the group individuals belong to (i.e., there is a higher risk for minorities like homosexual males) and on the tendency of others to disclose their sexual orientation.

Mislove et al. consider generic attribute inference and propose an approach based on social communities [Mislove et al. 2010]. They design an inference algorithm that first identifies communities that are centered on certain attributes (incl. gender and university) based on the structure of the social graph and the values of the attributes disclosed by some of the users (so-called “seeds”). The algorithm then uses these communities to infer the values of the attributes for the users who do not disclose them. The authors evaluate their approach on several datasets of a few hundred of thousand college/university students and focus on attributes related to student life, for example college, matriculation year (which is related to age), and major.

Thomas et al. also consider generic attribute inference and propose an approach that takes into account the strength of the social ties between users, based on the working assumption that the correlation between the attributes of two individuals increases with the strength of their social ties (which they confirm by analyzing a large dataset) [Thomas et al. 2010]. They use the number of common friends as a proxy for the strength of the social tie between two users.

¹⁶Because the authors had access to an archive containing multiple snapshots of the social network – i.e., Friendster – data, they were able to study the case of non-users based on the data that became available when non-users became users later on.

They rely on a logistic / linear regression-based classifier. They focus their evaluation on the following attributes: gender, political orientation, religious denomination, relationships status (e.g., single, married) and interests (e.g., movies, music); and they rely on two datasets of FB users in the US ($N \approx 43k$ and $41k$ respectively). Their results show that the information about the attributes of a users' friends substantially improves the accuracy of the inference process.

Pesce et al. propose a similar approach but use photo tags as a proxy to the strength of the social ties between users [Pesce et al. 2012]. They focus their evaluation on the following attributes: age, gender, city and country; and they rely on a dataset of $N = 664$ FB users in Brazil and India.

Kotyuk and Buttyan also consider generic attribute (numerical and nominal) inference, and they propose an approach that takes into account the correlation between the private attributes to be inferred [Kotyuk and Buttyan 2012]. They rely on a multi-layer perceptron (MLP); their approach can achieve both regression and classification. They focus their evaluation on the following attributes: age, gender and marital status; and they rely on a dataset of $N \approx 13k$ users of *iwu*,¹⁷ a popular Hungarian OSN.

Zamal et al. focus on the inference of three attributes covered by homophily: gender, age, and political affiliation [Zamal et al. 2012]. By relying on a support vector machine classifier and on a dataset of Twitter profiles, the authors show that (i) similar inference accuracy can be reached for age and political affiliation by using only friends' data instead of the target user's data, and (ii) by combining the user's and their friends' data, the inference accuracy substantially increases.

Dey et al. focus on age inference [Dey et al. 2012]. More specifically, they design an algorithm to infer the birth year of an OSN user, based on those of their friends and the friends of their friends (and so on). The authors design an ad-hoc iterative inference algorithm that propagates the information about a user's birth year to their friends by using regression techniques. They evaluate the performance of the proposed inference algorithms by using a sample of $N \approx 1.7M$ FB profiles from the NYC network. Their results show that age can be inferred with an accuracy of less than 2 years for more than 70% of the users. In their follow-up work, Dey et al. focus on the case of high-school students and demonstrate that, by using similar inference techniques and datasets, information hidden by default on the profiles of minors (e.g., birth year) can be inferred from the users' friends' profile information [Dey et al. 2013].

Ryu et al. analyze the inference of hidden attributes based on the local network of users in OSNs [Ryu et al. 2013]. To do so, the authors use three different algorithms and test them on three datasets: Google+ data, UCI FB data, and Duke FB data. The results show that the algorithms have similar performance for the different datasets and, for 75% of users, the algorithms perform much better than random guessing when inferring binary attributes.

Gong et al. perform both link prediction and user attribute inference, and they show that attribute inference can further improve link prediction, and conversely, that link prediction can improve attribute inference due to homophily [Gong et al. 2014]. The authors demonstrate the superiority of their approach compared to previous work by using Google+ data and by relying on various unsupervised and supervised algorithms, such as support vector machine.

Gong and Liu propose to combine friend-based attacks and behavior-based attacks to improve upon previous approaches on attribute inference [Gong and Liu 2016]. They focus, in particular, on the inference of three attributes: employer, cities where users have lived, and major. By relying on Google+ and Google Play data with $N \approx 1.1M$ user profiles, the authors show that their attack's precision outperforms friend-based attacks and behavior-based attacks by 20% and 100%, respectively. Jia et al. further improve upon this attack, both in terms of inference accuracy and algorithmic efficiency, by relying on Markov random fields and loopy belief propagation [Jia et al. 2017]. They evaluate the performance of the attack on a dataset of $N \approx 5.7M$ Google+/Google Play users.

¹⁷www.iwu.hu, last visited: Sep. 2018.

Alsarkal et al. focus on re-identification attacks based on data disclosed by others [Alsarkal et al. 2018]. More specifically, they investigate to which extent an individual can be re-identified, i.e., can be linked to an identity in the real world, based on this information about them revealed by other individuals (i.e., “co-disclosed” according to the terminology used by the authors). They rely on Shannon’s entropy – measured in bits – to capture the re-identifiability of an individual. They use as a baseline the case where individuals disclose information only about themselves (i.e., “self-disclosure”). Using a dataset of $N = 1,357$ Twitter user accounts, they demonstrate that an individual’s first name, age, and zipcode are significantly more often self-disclosed than co-disclosed and that an individual’s birthday and gender are significantly more co-disclosed than self-disclosed. They found no significant difference (between self- and co-disclosure) for family relationships. They also demonstrate that the privacy loss (in bits) caused by co-disclosure is substantial; up to two times more than that caused by self-disclosure in some cases.

As part of a study of data-collection practices of FB third-party apps, Wang et al. identify an interdependent privacy risk related to OSN data [Wang et al. 2011b]. The risk comes from the fact that, on some OSNs, including FB, third-party apps can request not only access to a user’s data but also to that of their friends (e.g., profile information such as demographics). And only the consent of the user is required; not that of their friends, thus creating an interdependent privacy risk (note that FB’s permission system has changed significantly since this work was conducted). Their analysis of a dataset of FB applications shows that 148 out of 1305 apps (11.3%) ask for friends’ personal information. This problem was also noted by four participants (out of 11 from an undergrad students population in the US) during interviews, stating “[the user’s] friends never download or agreed to the application’s terms” and “[the user] does not own [his or her] friend’s information”. Biczók and Chia further investigate this problem [Biczók and Chia 2013]. Their analysis of a dataset of FB apps shows that 518 out of about 27k apps (1.92%) ask for friends’ personal information, and that 18k applications (67%) request for basic information of user of the app; this information reveals also her list of friends. This specific interdependent privacy risk was recently put under the spotlight with the Cambridge Analytica scandal [Cadwalladr and Graham-Harrison 2018].

Humbert et al. study to which extent a social network user can be found through others’ friend lists due the fact that hiding their own friend list does not prevent their profile to appear in their friends’ friend list [Humbert et al. 2013b]. In particular, the authors propose a new navigation attack to discover users (even those who do not appear in the search directory) by relying on some known attributes about them (such as place of residence, workplace, or alma mater). They demonstrate through experiments on Facebook and Google+, that the majority of users can be found by crawling a median number of user profiles smaller than 400 and 300 respectively.

Concerns. Many works study the privacy concerns related to interdependent privacy risks on OSNs, yet, most of them focus on the risks related to photo sharing rather than on demographics and preferences. Therefore, we survey them in Sec. 4.3. Also, Pu and Grossklags study the value users put on the privacy of their friends, in the context of information disclosure on OSNs and access to friends’ data by third-party apps on OSNs [Pu and Grossklags 2014, 2015, 2017]; this is discussed in detail in Sec. 5.1.1 in the context of game theory models for non-cooperative solutions.

4.2 Genomic Data

Context. The genetic information of a human being (i.e., their genome) is encoded as a sequence of around 3 billion nucleotides (A, T, C, G) pairs [Klug et al. 2003]. In each pair of nucleotides in an individual’s genome, one nucleotide is inherited from their mother, taken at random from the mother’s corresponding nucleotide pair during the production of one of her reproductive cells, and the other one from their father, also taken at random during the production of his

reproductive cells. This rule is known as Mendel’s first law of inheritance (segregation of genes). As a consequence of genetic inheritance, the genomic data of an individual is correlated with those of their parents and children but also, by extension, with those of their family members at large.

Recent advances in genomics have enabled the development of direct-to-consumer genetic sequencing services (e.g., 23andme¹⁸) that enable individuals to obtain their genomic data from a biological sample (typically saliva). Genetic data collected this way is sometimes shared on online platforms (e.g., openSNP¹⁹) for various purposes, including finding relatives, learning about family history (origins), or advancing (medical) research. Therefore, such data becomes available to different entities, beyond the sequencing services, such as online platform operators or even the public. Not only is the genome of an individual immutable (it does not change over their lifespan) but it reveals very sensitive information about the individual, including their physical appearance and their predisposition to diseases (e.g., Alzheimer’s). Such information can be used for all sorts of discrimination, for instance when setting (health) insurance premiums.

Risks. Humbert et al. were the first to formalize the problem of interdependent privacy risks with genomic data between family members [Humbert et al. 2013a], also known as the *kin genomic privacy* problem (a concept introduced in [Stajano et al. 2008]). They rely on factor graphs, a particular type of PGMs (see Sec. 3.1), to model the statistical relationships between the genomic data of family members. Using this model, they conduct inference by using the belief propagation algorithm and by quantifying the privacy of family members, based on the output of the inference. In order to validate their model and assess the magnitude of the risk, they conduct data-driven simulations based on a CEPH/Utah Pedigree, a dataset consisting of the (partial) genomes of $N = 17$ family members (4 grandparents, 2 parents and 11 children). Their results show that the genome of an individual can be inferred to a large extent from those of their family members: In the considered dataset, the genomic privacy of the father (quantified as the inference error) is decreased by around 90% when the genomes of his parents, his wife, and four of his children are known. In their follow-up work, Humbert et al. further improve the inference attack by using a generic Bayesian network (see Sec. 3.1) and by taking into account phenotypic information (e.g., diseases or physical traits), hence showing that the risks are even more serious [Humbert et al. 2017]. Deznabi et al. improve the model of Humbert et al. by considering phenotypic information and high-order correlations between the genomic variants [Deznabi et al. 2018].²⁰ The authors run extensive experiments and demonstrate that high-order correlations can significantly improve inference, hence damage privacy, even with very few relative(s) sharing their genomic data. In a follow-up work, Humbert et al. propose an algorithm for evaluating kin genomic privacy in a data-less fashion, i.e., they rely only on the family tree and on the list of family members whose genomes are known to the adversary (not the genomic data itself) [Humbert et al. 2019]. To do so, they rely on Bayesian networks and consider, for each position in the genome, all possible combinations of nucleotides for the sequenced family members. The authors provide an online tool, in the form of a web application, that enables users to build their family trees and to evaluate the genomic privacy of their relatives (including themselves).²¹

Backes et al. study the impact of such genetic interdependencies at scale with the (synthetic) genotypes of $N = 1,000$ individuals over five generations [Backes et al. 2018]. These genotypes were generated through simulations. They evaluate to which extent the sharing of genomic data of a fraction of individuals in this population influences the privacy of others. They observe that the global genomic privacy decreases super-linearly w.r.t. the sharing rate.

¹⁸<https://23andme.com>, last visited: Feb. 2019

¹⁹<https://opensnp.org>, last visited: Feb. 2019

²⁰A genomic variant denotes a position in the human genome where the nucleotides differ in the population (i.e., not all human beings have the same nucleotide at this position). In the biomedical community, this is more precisely referred to as *single-nucleotide variant* (SNV).

²¹<https://santeperso.unil.ch/privacy/?survey>, last visited: Jun. 2019.

Berrang et al. generalize this probabilistic approach and propose a Bayesian network model that encompasses three dimensions of dependencies: (i) between relatives, (ii) between data at different points in time, and (iii) between different biomedical data types [Berrang et al. 2018]. More precisely, the authors rely on a dataset containing the genomic and epigenomic (DNA methylation) data of mother-child pairs, and at two or three different time points for some of the individuals. Unlike in previous work, the structure of the Bayesian network and its parameters are not fully known to the adversary; instead, they are learnt from training data by relying on maximum likelihood estimation (MLE) The results show that, by knowing the methylation data of the child and genomic data of the mother, the estimation error on the mother’s methylation data is reduced to 0.1 or less for 60% of the positions, compared to 10% of the positions when using only population statistics.

Ayday and Humbert survey the various inference attacks and possible solutions in kin genomic privacy [Ayday and Humbert 2017]. In addition to (some of) the aforementioned literature, they cover the works on membership inference in genomic databases. Membership inference consists in determining whether a target is part of a database by knowing some of their raw (genomic) data and comparing it to aggregate statistics. In this context, Sankararaman et al. show that it is possible to infer the membership of a first-degree relative (i.e., parents, children, or siblings) of the target with the same success as inferring the target’s membership by having access to four times more of genomic variants [Sankararaman et al. 2009]. Shringarpure and Bustamante present a novel membership inference attack against genetic data-sharing beacons [Shringarpure and Bustamante 2015]. A beacon is a web server that replies to queries about the presence of a given nucleotide at a certain position (in the human genome) in its database with binary answers (i.e., “yes” or “no”). In this context, the authors study the power of membership inference for various degrees of relatedness and show that it is possible to achieve high success for first-degree relatives by having access to 40k variants. This specific membership inference risk was recently put under the spotlight with the Golden State Killer case, where distant relatives of one of the suspects (whose DNA was found on the crime scene) were identified using the GEDmatch database,²² which eventually enabled the police to uniquely identify the culprit [Abrams 2018; Murphy 2018]. Note that, in this survey, we focus only on membership inference attacks that affect individuals because of *correlations*.

Concerns. Weidman et al. perform a vignette survey study ([Aviram 2012]) to explore the impact of key factors, such as demographic characteristics, on trust, personal and interdependent privacy, and on genetic data-sharing intentions [Weidman et al. 2018]. Specifically, the authors design two vignettes: (i) They ask the respondent to answer questions regarding the sharing of their own genomic data, and (ii) they ask the respondent to answer questions assuming friends or siblings seek advice about sharing their genomic data. The survey ($N = 524$ Amazon mechanical turk (MTurk) users) notably shows that if the genomic data is not anonymized (i.e., linked to an identity in the real world through a name), it is less likely that the owner of the genome shares this data or advises others to share their data. Furthermore, it shows that concerns for a relative’s privacy is significantly influenced by the entity requesting the data (government research group, private medical group, or academic research group) in the first vignette, and by the age of the advisor in the second vignette. In the same vein, De Cristofaro studies users’ perception of privacy and related ethical issues in an interview-based study involving $N = 16$ participants [De Cristofaro 2013]. In particular, this study demonstrates that, for 14 out of the 16 respondents, the case where siblings share their genomes for research induces less discomfort than the case of sharing labor or health insurance discrimination or of hacked genome.

²²<https://www.gedmatch.com>, last visited: Jun. 2019.

4.3 Multimedia Data

Context. With mobile devices, including most modern smartphones and tablets, individuals can capture everyday life situations in the form of photos and videos (i.e., multimedia content). Such contents often feature individuals other than the individual who captured them, including family members, friends, friends of friends, colleagues, and strangers. And this content is often shared on OSNs (e.g., FB) or dedicated platforms (e.g., Flickr²³), possibly with (face) tags identifying the individuals featured in them.

Unlike for other data types, with photos and videos, the data can be directly accessed by an attacker, hence he does not need to conduct any statistical inference. As such, photos and videos are multiple-subject data, not interdependent data (as defined in Sec. 2) per se. Therefore, the interdependent privacy risks associated with this type of data are mostly limited to the disclosure of the content itself. Yet, it is important to note that photos and videos (media) can contain sensitive information about the individuals featured in them, including various personal and contextual attributes correlated with – thus inferable from – an individual’s image or speech (e.g., gender and age [Levi and Hassncr 2015], ethnicity [Fu et al. 2014], emotion, activity), and the fact that the individuals featured in the media are together, which reveals some information about their relationship [Shoshitaishvili et al. 2015]. In some cases, multimedia contents can also leak the location, and more generally the context, of the individuals featured in them.

Risks. Raynes-Goldie studies the privacy risks on OSNs and the associated concerns through an ethnographic study of FB users ($N \approx 20$) in Toronto [Raynes-Goldie 2010]. The study tackles (among other things) interdependent privacy risks related to the fact that users can pass forward to others the content they have access to (e.g., photos). They mention, for instance, the case where a user blocked by an ex-lover asks a common friend to access and pass to the user the photos shared by their ex. Although the threat is real, it is very hard – if not impossible – to mitigate. In addition, the fact that the process is not automated reduces the extent of the threat.

Akcora et al. identify and investigate the case where the privacy policies and photo-tagging actions of users of OSNs put the privacy of their friends at risk [Akcora et al. 2012]. Indeed, when users set the visibility of the content they post to “friends only”, the content could actually be, in certain circumstances, visible to users other than their friends.²⁴ For instance, on FB, if a user comments on some content posted by one of their friends, the content becomes visible to their own friends, thus propagating beyond what the user who posted it intended.²⁵ The authors study the risk perception of users, toward the disclosure of photos to friends of friends and strangers, and the underlying reasons and factors.

Damen et al. also investigate the loss of control of users on the access to the content they or their friends share [Damen et al. 2014]. The authors survey such risks on FB and propose tools for assisting FB users in determining, depending on the actions of other users and the visibility of certain content including pictures.

Henne et al. identify and investigate issues related to user (un)awareness about the media (including photos) that features them on OSNs [Henne et al. 2013]. To do so, they rely on a survey of $N = 414$ users. The survey results highlight that 52% of the respondents learned by chance about shared photos that feature them. In some cases, when new content is shared by a stranger, the content cannot be removed because the featured user is not even aware of this content. Ilia et al. study the same issues, focusing on pictures, and they investigate the tagging habits of individuals [Ilia et al. 2015]. The authors explore this issue by collecting photos through a user study of $N = 128$ FB users. The results show that 92% of users have access to photos that depict strangers. They also show that individuals are not always tagged, as more

²³<https://www.flickr.com>, last visited: Feb. 2019

²⁴Yu et al. provide a comprehensive survey and analysis of the (default) privacy policies of several OSNs including FB and WeChat [Yu et al. 2018].

²⁵The visibility of comments now depends on the policy of the post, and can be restricted to friends and people tagged.

than 87% of pictures contain only one tag, and 75% contain at least two distinguishable faces. Such et al. focus on the same issues and show that the vast majority of users experience such multi-party privacy conflicts (99.3%) [Such et al. 2017], both as the users sharing the photos (uploader) and as those featured in the photos (co-owner). They also find that most of these conflicts are related to media captured during social events (e.g., Christmas) and happened on FB.

Concerns. Besmer and Richter Lipford study user privacy concerns related to the sharing and tagging of photos on OSNs [Besmer and Richter Lipford 2010]. Using three focus groups, with a total of $N = 14$ individuals, they identify the different actors who are perceived as adversaries (i.e., potential viewers) by individuals. Their results show that the privacy concerns are related mostly to individuals from personal social circles such as family members, friends, and employers, but not from strangers or from the service provider and its associates – a finding common to other studies, e.g., [Debatin et al. 2009; Raynes-Goldie 2010]. They also show that privacy concerns are caused by the fact that the photos portray non-normal behaviors, such as illegal activities or unflattering behaviors.

Wang et al. study general users regrets regarding post on OSNs, including in interdependent privacy situations [Wang et al. 2011a]. The study is based on data collected through surveys on MTurk ($N = 813$) and interviews ($N = 19$) from Facebook users recruited on Craigslist. The authors reported that three interviewees (out of 19) showed regrets about sharing group photos (e.g., at parties) that impacted others negatively, showing them with embarrassing behaviors (e.g., underage drinking), or with an unflattering look.

Choi et al. study user perception regarding the disclosure, by others, of embarrassing contents such as photos on OSNs [Choi et al. 2015; Choi and Jiang 2013]. The authors rely on the CPM theory (see Sec. 3.4). In particular, building on an abundant literature for the offline case (i.e., in the real world, by opposition to the Internet), they focus on the situations where some content is disclosed on an OSN in order to tease a target individual. To test their model, the authors conducted a user study with $N = 109$ FB users from a student population of a university in south-east Asia; and they relied on hypothetical embarrassing disclosures. The authors investigate which factors influence the target individuals' perception of the privacy invasion and show that, in the general case, tagging the content with the identity of the target increases the perceived invasion. They also study how target individuals react to such invasions and how the perceived invasion influences their reactions.

Chen et al. focus on privacy concerns when content is disclosed by peers; more specifically, they investigate how individuals would make decisions if they could control – to some extent – which content about themselves others could disclose on OSNs [Chen et al. 2015]. They put an emphasis on photos, and more specifically on photo tagging. They introduce the concept of information privacy concern about peer disclosure (IPCPD) and rely on CPM and impression management to analyze such privacy concerns. They focus on two aspects of privacy management theory: the “what” (i.e., the content) and the “for whom” (i.e., the audience). And they capture these aspects through the discrepancy between the image projected by the content and the online identity of the individual, and the social network overlap (between the individual featured in the content and the individual who shares it). To test their model, they conduct a laboratory experience with $N = 139$ undergraduate students, by using short scenarios mixing high and low values for the three parameters of the model (i.e., discrepancy, social network overlap and decisional control), where students had to evaluate their privacy concerns and their control variables. The results show that privacy concerns depend on the type of content to be disclosed and on the audience who has access to the content. More importantly, the privacy concerns change in function of the degree of decision control that users have.

Ozdemir et al. also study the privacy concerns of individuals when content is disclosed by peers (which they call disclosure in peer contexts) [Ozdemir et al. 2017]. The authors rely on the “antecedents, privacy concerns, outcomes”

(APCO) model to evaluate these concerns, and extend it with the constructs of risk, trust, and benefit. They define several hypotheses to test their model, e.g., higher privacy awareness is associated with higher information privacy concern, higher risk is associated with lower perceptions of trust, and more perceived benefit increase the information disclosed online. Then they evaluate the hypotheses with an online survey questionnaire $N = 314$ US FB users. Their results support their hypothesis and highlight the importance of privacy experiences and the privacy awareness construct in individuals' privacy concerns. They also show that information disclosure significantly depends on the benefits obtained by sharing content, the trust in other peers and the privacy concerns of individuals.

Symeonidis et al. study user concerns regarding the interdependent-privacy risks of third-party social network apps (identified by Wang et al. and investigated by Biczók and Chia) that they refer to as collateral damage [Symeonidis et al. 2016a,b]. They rely on an online survey with $N = 114$ FB users. They find that 66% of the participants are at least very concerned about the fact that, by default, FB enables apps to access information about a user's friends; and 77% are at least very concerned about not being notified when such situations occur. Furthermore, they find that users are also concerned about the privacy risks they expose their friends to, thus showing some form of altruism regarding privacy, as considered in other works (e.g., [Pu and Grossklags 2014, 2015, 2016]). We discuss these works in Sec. 5.1.1. In a follow-up work, Symeonidis et al. revisit their initial investigation in the light of the Cambridge Analytica scandal and complement it with a legal analysis based on the General Data Protection Regulation of the European Union (EU GDPR) [Symeonidis et al. 2018].

Jia and Xu study collective privacy concerns in OSNs, which rely on the CPM theory [Jia and Xu 2016b] (see Sec. 3.4). The authors evaluate these concerns at diffusion, access and control dimensions. To do so, they study the current literature about privacy concerns and they collected data via an online survey, from the US population of undergraduate students ($N = 427$). Their results show that individuals are concerned with these three dimensions of privacy and more than 50% of participants are frequent users of OSNs, hence the importance of these new privacy concerns.

Chutikulrunsee and Burmeiste study Facebook users privacy concerns (including conflicts) for photos sharing in OSNs [Chutikulrunsee and Burmeiste 2017]. The authors collected data about user behaviors regarding shared contents (i.e., photos, tags, comments) deletion requests, with an online survey with $N = 460$ participants recruited on SurveyMonkey. The results show that more than one-fourth of the participants requested (at least once) another user to remove a shared picture in the past.

4.4 Location Data

Context. Mobile devices, including most modern smartphones, can determine their current locations through a number of technologies and techniques, including GPS, Wi-Fi, and IP geolocation. Location information enables a variety of so-called location-based services. For instance, individuals can query dedicated services for the list of neighboring points of interest. They can also check-in at specific locations and venues or upload geo-tagged photos on OSNs. Location data is also collected by online services (typically through mobile applications) for advertising purposes. Therefore, location data becomes available to different entities. And this data is highly sensitive, beyond providing the ability to track the location of an individual: Research shows that sensitive personal information, such as demographics and sexual and political orientation, can be inferred from an individual's location data (e.g., [Zhong et al. 2015]).

When co-location data (i.e., data indicating a relationship between the locations of two or more individuals) is available, the location data of an individual leaks information about the location of other individuals. For instance, if co-location data indicates that two individuals are together – therefore their locations are the same – the location data

of one of them reveals the location of the other. Sources of co-location data include posts (with tags, e.g., “with John”) and photos (with faces tagged or identity information embedded in the meta-data of the photo) on OSNs.

Attacks. Sadilek et al. tackle the problem of joint inference of friendship and location for social network users [Sadilek et al. 2012]. More specifically, the authors present an attack that infers friendships between users based on (among other things) co-location profiles and uses the inferred friendship information to later infer the location of a user based on that of their friends. The proposed attack relies on a simple temporal graphical model (one per user; the locations of the users are not inferred jointly) similar to a HMM (see Sec. 3.1), and it makes use of a variant of the forward-backward algorithm. They evaluate the attack on a dataset of $N \approx 1.2M$ Twitter users based in the US (New York City and Los Angeles) and show that it can accurately infer the location of a user, based on that of their friends.

Henne et al. study interdependent privacy risks, based on the content and meta-data of photos uploaded on online platforms such as OSNs [Henne et al. 2013] (discussed in Sec. 4.3). As the meta-data of photos can include both location and identity information, the risk described in this work falls in the category of location privacy as well.

Jurgens studies location inference in OSN based on relationship between users, using a label propagation algorithm [Jurgens 2013] that propagates location information about a user to their contacts. The authors collected data from Foursquare and Twitter to evaluate (cross-platform) inference based on prediction linguistic similarity, self-reported locations and also timezone boundaries (to filter out incorrect distant locations). Their results show that it is possible to infer the location of 50% of the users with an error of less than 10km.

Vratonjic et al. identify an interdependent location-privacy risk, based on the use of shared public IP addresses [Vratonjic et al. 2013]. As users connected to an access point that uses network address translation (NAT), a common technique for routers deployed in homes and public places, have the same public IP address, the fact that two individuals have the same IP address constitutes co-location data: It indicates that they are likely at the same location (e.g., within the communication range of the access point). If an adversary learns the location of a single individual connected to the access point, it learns the location of the access point and those of all the other individuals connected to it (identified through their common public IP address that is known to the online service they use). The authors propose a theoretical probabilistic framework to quantify the risk, based on a number of characteristics of the Internet traffic of an access point. They also experimentally quantify the extent of the risk by using traffic logs from a university campus Wi-Fi network. In their follow-up work, Vratonjic et al. refine their theoretical model to account for the diurnal traffic pattern and demonstrate that it fits reality [Vratonjic et al. 2014].

Olteanu et al. were the first to identify and to formalize the general problem of interdependent location-privacy [Olteanu et al. 2014]. They focus on the cases of posts and photos uploaded on OSNs and tagged with the names of multiple individuals. They extend the state-of-the-art theoretical framework to quantify location privacy by including observed co-location data. Their model is based on HMMs (see Sec. 3.1) and includes location and co-location reporting (in the presence of obfuscation and errors). For increased accuracy, they adapt the original inference algorithm (i.e., forward-backward algorithm) to exploit co-location data. They show that, with co-location data, the complexity of the inference process increases exponentially with the number of target individuals, and they propose a heuristic of reasonable complexity. They quantify the extent of the risk experimentally by relying on a dataset of location traces (i.e., the GeoLife dataset), that they enrich with synthetic co-location data. Their results show that co-location data substantially decreases the location privacy of individuals. In their follow-up work [Olteanu et al. 2017], they refine their model by relying on Bayesian networks, which enables them to scale the inference process through belief propagation. They also take into account the case of erroneous co-location data.

4.5 Aggregate Data and Differential Privacy

The correlation between the data of different individuals raises issues not only for individual data (usually through inference attacks as described in the previous subsections) but also for aggregate data. In this setting, the now well-established approach for guaranteeing individual data contributors' privacy is to add some noise on the released aggregate data to achieve differential privacy (DP) [Dwork et al. 2006]. This definition guarantees that the distribution of any aggregate query result changes only slightly when a single record in the dataset is added/removed. However, for guaranteeing a certain privacy level, DP assumes that the different records in the considered dataset are statistically independent. Kifer and Machanavajjhala are the first to demonstrate, based on illustrative examples and synthetic data, that correlated records in a dataset can void the guarantees provided by DP [Kifer and Machanavajjhala 2011]. Liu et al. further demonstrate, based on real data, such correlation can be exploited through inference attacks and therefore jeopardize the DP guarantees [Liu et al. 2016]. Specifically, they rely on a dataset from Gowalla, composed of $\sim 100k$ check-ins from $\sim 7k$ users with $\sim 47.5k$ edges between them. The considered setting is that of a data provider that releases the centroids of clusters μ (computed with k -means from the users' check-in locations) in a differentially-private manner by applying Laplace noise on the original centroids. The adversary is assumed (i) to have access to all but the target user's data (a generic assumption in differential privacy), and (ii) to know the joint distribution between the data records in the dataset. The second assumption makes the adversary more powerful than the traditional DP adversary, but also more realistic when data records happen to be correlated (such as in OSNs). The authors first show that the adversary's posterior on the estimated value of the target record \hat{D}_i is as follows:

$$P(\hat{D}_i = \hat{\mathbf{d}}_i | \tilde{\mu}, \mathbb{D}_{-i}) \sim \exp(-|\tilde{\mu} - \hat{\mu}| \epsilon) \cdot P(\hat{D}_i = \hat{\mathbf{d}}_i | \mathbb{D}_{-i}), \quad (4)$$

where $\exp(-|\tilde{\mu} - \hat{\mu}| \epsilon)$ is the noise induced by the Laplace distribution applied on the centroids. Under the original DP adversary, $P(\hat{D}_i = \hat{\mathbf{d}}_i | \mathbb{D}_{-i}) = P(\hat{D}_i = \hat{\mathbf{d}}_i)$, which limits the power of the adversary to infer the posterior of \hat{D}_i . In this case, \mathbb{D}_{-i} can only be used to estimate the prior probability $P(\hat{D}_i = \hat{\mathbf{d}}_i)$. These records can indeed be used to sample D_i , typically by counting the number of check-ins falling into a given region. Now, if the adversary not only knows location records \mathbb{D}_{-i} but also some dependencies between D_i and these (e.g., through social relationships of users), $P(\hat{D}_i = \hat{\mathbf{d}}_i | \mathbb{D}_{-i})$ can be better estimated by assigning higher weights to the check-ins whose owners are friends with the target (due to the fact that mobility patterns between are often correlated). Finally, the authors show that inference attack achieve better performance when accounting for dependencies and that the theoretical maximum information leakage guaranteed by DP is violated under a more realistic adversary who exploit correlation between records.

5 SOLUTIONS

In this section, we survey the different solutions for protecting against interdependent privacy risks presented in the literature. We distinguish between non-technical (based on individual decision making and simple actions, with or without cooperation between the involved individuals) and technical (based on advanced techniques such as cryptography and access control²⁶ solutions). Note that we do not cover the solutions put in place by service operators (e.g., OSN operators). Such operators have indeed put some effort in this direction over the past years, especially in the case of revenge pornography. Yet, such solutions do not protect user privacy with respect to the service providers, and they often fail to protect against other users due to the deletion delays [Liang et al. 2015]: the damage is done.

²⁶Paci et al. wrote a comprehensive survey on access control solutions for collaborative systems [Paci et al. 2018].

Table 2. Summary of interdependent privacy risks and associated attacks. For the adversary, the platform operator is abbreviated “op.” and the public “pub.”. The different models and inference algorithms are abbreviated as follows: Bayesian network (BN), belief propagation (BP), forward-backward (FB), factor graph (FG), hidden Markov model (HMM), junction tree (JT), label propagation (LB), logistic/linear regression (LR), Markov random fields (MRF), multi-layer perceptron (MLP), naïve Bayes (NB), random forest (RF), support vector machines (SVM).

article	main data	auxiliary data	platform	adversary	inference
[He et al. 2006]	profile (all)	social ties	social networks	op., pub.	BN
[Becker and Chen 2009]	profile (all)	social ties	social networks	op., pub.	voting
[Lindamood et al. 2009]	profile (political)	social ties	social networks	op., pub.	NB
[Zheleva and Getoor 2009]	profile (political, gender)	social ties/groups	social networks	op., pub.	ad-hoc
[Jernigan and Mistree 2009]	profile (sexual)	social ties	social networks	op., pub.	LR
[Sarigol et al. 2014]	profile (sexual)	social ties	social networks	op., pub.	RF
[Misllove et al. 2010]	profile (political, gender...)	social ties	social networks	op., pub.	LR
[Thomas et al. 2010]	profile (all)	social ties	social networks	op., pub.	ad-hoc
[Kotlyuk and Buttyan 2012]	profile (all)	social ties	social networks	op., pub.	MLP
[Zamal et al. 2012]	profile (political, gender...)	social ties	social networks	op., pub.	SVM
[Dey et al. 2013, 2012]	profile (age)	social ties	social networks	op., pub.	LR
[Ryu et al. 2013]	profile (all)	social ties	social networks	op., pub.	LR
[Gong and Liu 2016]	profile (employer, location...)	social ties	social networks	op., pub.	ad-hoc
[Jia et al. 2017]	profile (employer, location...)	social ties	social networks	op., pub.	MRF+BP
[Alsarkal et al. 2018]	profile (gender, age...)	N/A	social networks	op., pub.	direct
[Wang et al. 2011b]	profile (all)	N/A	social networks	3 rd -party app	direct
[Biczók and Chia 2013]	profile (all)	N/A	social networks	3 rd -party app	direct
[Humbert et al. 2013a]	genomic	family ties	genomic/ancestry	op., pub.	FG+BP
[Humbert et al. 2017]	genomic	family ties	genomic/ancestry	op., pub.	BN+JT/BP
[Humbert et al. 2019]	genomic	family ties	genomic/ancestry	op., pub.	BN+JT/BP
[Backes et al. 2018]	genomic	family ties	genomic/ancestry	op., pub.	BN+JT/BP
[Berrang et al. 2018]	genomic and epigenomic	family ties	genomic/ancestry	op., pub.	BN+JT/BP
[Akcora et al. 2012]	photo	N/A	social networks	pub.	direct
[Damen et al. 2014]	generic posts (incl. photo)	N/A	social networks	pub.	direct
[Sadilek et al. 2012]	location	social ties	social networks	op., pub.	BN
[Henne et al. 2013]	photo, (co)-location	N/A	photo hosting	op., pub.	direct
[Jurgens 2013]	location	social-ties	social networks	pub.	LB
[Ilia et al. 2015]	photo	N/A	social networks	pub.	direct
[Such et al. 2017]	photo	N/A	social networks	pub.	direct
[Sadilek et al. 2012]	location	co-location/social ties	social networks	op., pub.	HMM+FB
[Vratonjic et al. 2013, 2014]	location	IP	service providers	op.	direct
[Olteanu et al. 2014]	location	co-location	social networks	op., pub.	HMM+FB
[Olteanu et al. 2017]	location	co-location	social networks	op., pub.	BN+BP

5.1 Non-technical Solutions

In this section, we focus on non-technical solutions, that is solutions that involve simple manual user actions such as sharing/unsharing (or setting the visibility of) some data and befriending/unfriending some users.

5.1.1 Non-cooperative Approaches. We consider here situations with interdependent privacy risks in which rational²⁷ individuals singlehandedly choose to adopt certain strategies in order to optimize their own risks and benefits. Due to the interdependent nature of the considered situations, the risks and benefits of the individuals also depend on the strategies adopted by the other individuals involved. As such, these individuals interact with each other (through a

²⁷Recent studies have shown that individuals behave rationally in some situations when making security/privacy-related decisions [Redmiles et al. 2018]

so-called *game*) but do not cooperate. We focus on solutions where the individuals eventually reach a state where no individual has an incentive to change their strategy. This type of setting is generally studied by relying on game theory (see Sec. 3.2) and, more specifically, on so-called non-cooperative games. Note that, orthogonal to this work, there are interdependent *security* games that have also been extensively studied; Laszka et al. wrote a comprehensive survey on the topic [Laszka et al. 2014]. This line of research and the associated terminology certainly influenced the naming of the first work on interdependent *privacy* games by Biczók and Chia. Prior to that, Manshaei et al. wrote a comprehensive survey on the use of game theory for studying security and privacy [Manshaei et al. 2013].







Squicciarini et al. were the first to address the problem of *privacy management* on generic shared (co-owned) data with a game-theoretic model [Squicciarini et al. 2009]. They rely on mechanism design and in particular the Clarke-Tax mechanism [Clarke 1971] to derive the privacy policy that maximizes the social utility of the individuals co-owning the data. The Clarke-Tax mechanism is a simple voting scheme for social choice on a common good (such as the co-owned data) and has a notable advantage of being not manipulable by individuals and to ensure truthfulness. In this setting, individuals have to elicit their privacy preferences for every new shared item. In order to free the individuals from the burden of setting these preferences for each of the co-owned item, Squicciarini et al. further propose an inference algorithm based on previous sharing decisions. The authors propose and implement a proof-of-concept FB application for collective privacy policies enforcement, and show that the execution time of the Clarke-Tax mechanism increases approximately linearly with the number of co-owners.

As discussed in Sec. 4.1, third-party applications on OSNs can, in some cases, access the data of the friends of the user who installed them, thus creating interdependent privacy risks. Biczók and Chia investigate such risks with third-party FB app permissions [Biczók and Chia 2013]. They define a non-cooperative game between two FB users (choosing between install and *not* install an app), and analyze how positive and negative externalities (including privacy) influence the adoption of an application. They show that negative externalities can lead to equilibrium outcomes that are inefficient for both users and for the service provider. Finally, they discuss the possible causes of inefficient outcome, such as misaligned incentives, the absence of risk signals, and the low user awareness of interdependent privacy risks.

One limitation of this first approach is that, as it is limited to two users, it does not take entirely into account the dynamics of application adoption in today's OSNs and mobile networks. Pu and Grossklags fill this gap by simulating the application adoption process on scale-free networks [Pu and Grossklags 2014]. They build upon Sundararajan's local network effects model [Sundararajan 2008] to define a payoff function that takes into account (i) the direct value of using the application, (ii) the positive network effects, (iii) the privacy harm inflicted on friends when installing an application (this falls in the scope of altruism as defined by Meier et al. [Meier et al. 2014]), and (iv) other installation costs. Their results show that decreasing the installation cost or the interdependent privacy harm increases the early application adoption. They also indicate that app adoption increases as the network size increases.

In order to quantify the (monetary) value of interdependent privacy, Pu and Grossklags rely on conjoint analysis, i.e., to measure users' preferences in a trade-off setting [Pu and Grossklags 2015]. The results of their user study (based on $N = 400$ MTurk users) and analysis first show that friends' privacy is the third most important application feature, after the price and the user's own privacy. However, when quantifying the monetary value OSN users associate with their own profiles and their friends' profiles, it can be observed that the value of an average friend's information is a tiny fraction of the value of a user's personal information. In fact, the total value of all friends' profiles is lower than the value associated with their own profile. In a follow-up work, the authors evaluate the influence of users' past privacy invasion experiences, privacy knowledge, trust on applications' data practices, online social capital, and privacy concerns related to the adoption of an application [Pu and Grossklags 2016]. They make use of structural equation

Table 3. Game-theoretic approaches for studying interdependent privacy problems.

article	data	game type	# of players
[Squicciarini et al. 2009]	 generic	Clarke tax mechanism	N
[Biczók and Chia 2013]	 profile (OSN)	simultaneous	2
[Pu and Grossklags 2014]	 profile (OSN)	network	N
[Hu et al. 2014]	 generic	repeated	N
[Humbert et al. 2015]	 genomic	simultaneous	N
[Olteanu et al. 2019]	 (co-)location	sequential	2

modeling (SEM) analysis ([Gefen et al. 2000]) to evaluate the effect of these factors on users' valuations of their own privacy and of their friends'. Among other insights, their user study (based on $N = 397$ MTurk users) shows that having encountered privacy invasion in the past has a negative effect on users' trust in the application's data management practice, which also significantly affects users' concerns for their own privacy. In order to address low data quality, Pu and Grossklags further improve upon their prior work ([Pu and Grossklags 2016]) by making use of choice-based conjoint analysis [Pu and Grossklags 2017]. Moreover, they consider four novel treatment conditions in two different dimensions: sharing anonymity (Are friends' shared data anonymous or not?) and context relevance (Are the requested friends' data useful to application functionality or not?). Their survey (based on $N = 931$ MTurk users) shows that treatment conditions affect the valuation of interdependent privacy: (i) Users tend to value significantly less their friends' data if they believe it is anonymous, and (ii) they also tend to value their friends' data less if they believe this data is useful for the application functionality.

Several works propose unfriending (i.e., removing a previously declared social tie on an OSN) as a solution for addressing interdependent privacy risks related to OSN profile information inference and leaks (discussed in Sec. 4.1) [Becker and Chen 2009; Gundecha et al. 2011, 2014]. For instance, Gundecha et al. propose a solution for identifying friends who compromise the most a users' privacy on an OSN. We do not survey these works because unfriending offers only limited protection (even though the information is removed, the social network operator can retain and exploit it anyways), and because none of these works study the interplay between users when users unfriend.

As discussed in Sec. 4.2, genomic data is also subject to interdependent privacy risks because of the correlation between the genomes of relatives [Humbert et al. 2013a]. Humbert et al. analyze the interplay between members of the same family when making decisions related to genomic-data sharing and storage security [Humbert et al. 2015]. The authors first consider a two-player setting and, by relying on a Markov chain model, derive a closed-form expression of genomic privacy levels and of NE (see Sec. 3.2). They also take into account potential altruistic behavior between family members by adding to each individual's payoff function the (weighted) utility of other family members. Contrary to previous interdependent privacy or security games that rely on theoretical interdependence models, typically linear, here the interdependent risks are given by the familial genetic correlations. The authors also extend their analysis to N players by relying on multi-agent influence diagrams (MAIDs), a graphical model that not only includes random variable nodes (like Bayesian networks) but also decision and utility variable nodes (see Sec. 3.2). Their results notably show that, unless the players coordinate, altruism can lead to suboptimal social outcome compared to the purely selfish setting due to players being too cautious to not put other players' privacy at risk.

As discussed in Sec. 4.3, multimedia data shared on OSNs relates to more than one individual and to individuals who do not have any control over the shared content. This is typically the case in photos where multiple individuals can appear and be tagged by friends. In order to analyze the behavior of strategic OSN users, Hu et al. present a multiparty

control game which encompasses N OSN users who collaboratively control the sharing of a data item in the OSN based on their privacy and sharing preferences [Hu et al. 2014]. The authors first show that their game converges to a unique NE in a few iterations. Then, they perform two user studies to evaluate to which extent real users behave as expected by the game-theoretic model. These studies demonstrate that users believe that everyone appearing in a photo should have the right to decide who can view it and that a user's privacy sensitivity is highly related to the content of the photo. Interestingly, the survey also shows that users may not adopt the best strategies when making decisions, essentially because of some altruistic behavior with respect to others. Indeed, users may not always maximize their own benefits without taking into account others' and they often care more about others' privacy than their own sharing benefits.

As discussed in Sec. 4.4, location data can lead to interdependent privacy risks when the individuals also share co-locations online, i.e., mention that they are with other individuals. In order to study the interplay between individuals with possibly different views on data sharing and privacy, Olteanu et al. develop a two-player game-theoretic model that takes into account the benefits and privacy implications of location and co-location sharing in OSNs [Olteanu et al. 2019]. The authors rely on conjoint analysis to quantify the various parameters of the game-theoretic model. More precisely, based on a user survey of $N = 250$ FB users (from MTurk), they evaluate users' preferences between (i) sharing and viewing posts, (ii) location and co-location information, and (iii) location privacy and sharing benefits. The authors observe that there is no strong consensus for one preference vs. another. In particular, for privacy vs. sharing benefits, there are two clearly separated classes, where 63% of the participants favor privacy over social benefits and the rest benefits over privacy. By embedding these values into the payoff functions, Olteanu et al. show a user can have a strong incentive to share their locations once a co-location has been shared by the other user with them, thus creating a vicious circle of location over-sharing.

Table 3 summarizes the main literature covered here on game-theoretic models for interdependent privacy settings. We notice that there are various contexts and game types. Also, two models can deal with only 2 players while the two others can handle $N \geq 2$ players.

5.1.2 Cooperative Approaches. We focus here on solutions where individuals interact and collaborate together to progressively and collaboratively reach a decision regarding information sharing and access control rules. Unlike the solutions described above, individuals collaborate relying on communications to negotiate the sharing rules and the audience of content, in order to obtain commonly acceptable results. Individuals try to reach a common goal to build trustworthy relationships and improve their relations.

Besmer and Richter Lipford propose a corrective method as a collaborative access-management tool for photo sharing on FB: it enables co-owners (i.e., individuals who appear on the photo) to suggest restrictions to the owner (i.e., the uploader) [Besmer and Richter Lipford 2010]. The authors use three focus groups (with a total of $N = 14$) from a US university population (from staff and students) to identify the concerns of individuals and the coping mechanisms they use to limit the negative effects of shared photos. The results show that individuals are concerned about photos that could affect their social images. The main coping mechanisms used by co-owners are (1) individuals change their behavior to avoid appearing on undesired pictures, (2) they share through private channels (e.g., instant messaging and e-mails) to limit the audience and, (3) they interact outside of the OSN platform to negotiate photos sharing rules. The authors test their tool through an in-lab study ($N = 17$). Unlike untagging, restricting the access to the shared content requires that a request is explicitly sent to the uploader; for this reason, interviewees were less comfortable using the tool. Interviewees also reported that collaborative strategies improve long-term relationships, with benefits such as satisfaction, trust, and affection.

Lampinen et al. study preventive and corrective strategies, respectively, employed before and after the upload of a new content that creates privacy conflicts, used by individuals to regulate their privacy boundaries [Lampinen et al. 2011]. The authors collected qualitative data from interviews and focus groups (with a total of $N = 27$) from a student population in Finland. Their results show that interviewees employ both individual and collaborative strategies to manage access control. For preventive strategies, to block new publications, they employ the technique of separating audiences and rules of thumb (e.g., not make choices that do not match their own privacy rules). For corrective strategies, individuals use delete comments, untag, and ask the owner to delete content. The results show that individuals more easily use individual strategies, that do not require others' decisions.

Carminati and Ferrari present a global architecture for collaborative access control on any kind of data, including data shared between multiple OSN users [Carminati and Ferrari 2011]. For data shared between more than one user, the proposed system relies on a module that requests feedback from all users involved in the requested file, before granting others access to it (based on some predefined policy, e.g., majority vote).

Squicciarini et al. propose a collaborative privacy management tool, named CoPE: it enables OSN users featured in some shared content (referred to as co-owners) to determine which users of the social network can access this content [Squicciarini et al. 2011]. The authors focus and implement their solution for (group) pictures shared on FB. Simply stated, each co-owner simply specifies a set of users who can access the shared content; and the final access control list is obtained by intersecting the individual lists specified by the co-owners. The work sheds some light on user perception of such a tool: The authors implement a proof-of-concept of CoPE and conduct a survey of FB users ($N = 80$) from a student population in the US. The results show a high intention to use CoPE, regardless of the users' privacy concerns, as well as a high perceived usefulness, ease of use, and likability.

To reduce the risk of conflicts, Wisniewski et al. identify coping mechanisms that use privacy boundary regulation [Wisniewski et al. 2012]. The authors identify the different mechanisms by interviewing individuals ($N = 21$) recruited via e-mails and through snowball sampling on FB. Their results show that individuals mostly manage their boundaries by controlling the audience with the following methods: (1) filter friend requests, (2) ignore posts from other OSN users, (3) to prevent others from finding them, they use pseudonyms, (4) to avoid any conflicts, they withdrawal with self-censorship (e.g., interacting less with other users, not posting content), (5) to find compromise, they reach mutual sharing decisions and they communicate outside the OSN (e.g., with private communication). The results of the study highlight the lack of collaborative tools integrated into OSN platforms; this lack forces individuals to use side-channels to communicate such as private messages or face-to-face.

Marwick and Boyd focus on teenagers privacy management and on networked privacy (a concept similar to interdependent privacy; see Sec. 2) [Marwick and Boyd 2014]. The authors study the strategies used by interviewing and observing teenagers ($N = 166$) from the US population. Their results show that teenagers employ different coping mechanisms to control the access to the content shared online. The interviewees report that they manually filter content by sharing new content to different groups, and employ social steganography (e.g., talk about someone without explicitly tagging or mentioning them) to code their posts so that only friends who share a common context can understand the (hidden) meaning. They also rely on trust and respect to ensure that friends do not post embarrassing content.

Ratikan and Shikida introduce a collective privacy protection (CPP) tool for information shared on OSNs [Ratikan and Shikida 2014]. The CPP relies on majority voting: the owner proposes a privacy policy to all identified co-owners. Then, these co-owners accept or reject the policy (if no response is provided, the default choice is rejection to preserve privacy). The authors evaluate their solution with participants ($N = 24$) who observed the process as both roles (i.e.,

owner and co-owner) from scenarios relying on a virtual social graph and then answered a survey. The results show that CPP improved the users' perception of privacy protection against the leakage of shared information.

Cho and Filippova study the mechanisms used by FB users to control their (networked) privacy [Cho and Filippova 2016]. The data collection was done in two steps: to extract practices to manage networked privacy, they interviewed five focus-groups (with a total $N = 28$) from a student population in Singapore, and to evaluate the different practices found (e.g., collaborative, preventive and corrective strategies), they collected data from an online survey ($N = 299$) from a US and Singapore based population. The results show that individuals use corrective and preventive mechanisms, but also use information control mechanisms such as censorship. Collaborative strategies, however, such as discussing rules of thumb, discussing privacy settings, asking permissions and educating, are employed outside of OSN platforms. The authors also note that privacy behaviors are affected by privacy concerns and collective efficacy (i.e., the social cohesion among a group of individuals), and they show that the management of networked privacy is mostly a collective process.

Jia and Xu developed a theoretical framework of the collaborative privacy-management strategies of OSN users, based on the ownership of a content (e.g., photos), and on its access by individuals and the extension of the access (e.g., with repost) [Jia and Xu 2016a]. The authors base their work on the communication privacy-management theory and define boundary ownership (e.g., the co-owners of the content), permeability (i.e., the access rules) and linkage (i.e., the extension of the access) management. They collect their data from two online surveys from US based MTurk users ($N = 304$) and undergraduate students of a US university ($N = 427$). They find that collaboration and groups characteristics have an influence on the choice of privacy policies of individuals.

Such and Criado introduce an automated conflict resolution model based on concession, to help users manage the access control rules of content posted on OSNs [Such and Criado 2014]. Their model defines a set of individuals (co-owners who are tagged in the post) who collaboratively decide the access of a co-owned item (e.g., a group photo, managed by all the individuals who appears on the photo) for a set of target individuals (i.e., the audience of the photo) characterized by a certain type of relationship. Depending on the intimacy of relationships of the target individuals, they could have access if there is no conflict in the privacy rules. To resolve conflicts, the authors base their solution on co-owners' willingness to change their policy, depending on the sensitivity and the level of intimacy with the initial choice. In their follow-up work [Such and Criado 2016], the authors test their solution using a user study with $N = 50$ participants (from staff and students as well as volunteers outside academia) recruited by e-mail. Each participant had to specify their privacy policy in ten different fictional scenarios. They compared the changing behavior of users to their solution and to the main existing rules, such as veto voting (i.e., use the most restrictive rule), majority voting (i.e., match the maximum of co-owners' rule), and uploader choice (i.e., correspond to the actual rule in OSN, where only the uploader decides to share the content or not). Their results show that the proposed automated conflict-resolution technique outperforms the three generic rules, matching more than 80% of the concession behavior of users, followed by veto voting that always denies access to users hence is more restrictive than the developed model. Such and Criado list the actual techniques used to deal with multiparty privacy conflict (MPC²⁸), such as manual, auction-based, aggregation-based, adaptive, game-theoretical and fine-grained approaches [Such and Criado 2018]. Moreover, the authors state that the issue of the current research is the lack of large-scale real-world studies: to be able to compare different solutions, and to remove bias when participants do not recognize themselves into what-if scenarios used in surveys.

To manage the access to photos, Fogues et al. introduce a collaborative solution, based on the negotiation of co-owners via arguments [Fogues et al. 2015]. These arguments support each co-owner's choice of audience and are divided into

²⁸not to be confused with Multi-Party Computation

three categories: consequences (the impact of sharing the content), analogy from previous experiences, exceptional cases and popular opinion. In their follow-up work [Fogues et al. 2017a], the authors developed an inference model to test, with a survey of individuals from MTurk ($N = 988$), the influence of arguments of preferences and of context in the sharing of new content online. In five different scenarios, each participant had to specify a sharing policy, depending first on contextual factors, then with preferences and finally with the arguments associated with the preferences. Their model shows the importance of sensitivity and arguments in predicting the best sharing policy. Fogues et al. implement a recommending sharing policies agent in the case of multiparty content, which relies on the features considered in their previous work [Fogues et al. 2017a] and new features based on user and group characteristics (e.g., age, education level, past experience) [Fogues et al. 2017b]. The authors trained the recommending system by collecting data from MTurk participants; they did not specify the number of participants. Their results show that the proposed solution provides better accuracy compared to standard solutions, i.e., veto and majority voting.

Kamleitner and Mitchell list and study, based on a number of informal interviews with various individuals (incl. legal experts, scholars, and students), 24 typical cases that lead to interdependent privacy violations and conceptualize the problem through a framework [Kamleitner and Mitchell 2019]. In their legal analysis, focused on the EU GDPR, the authors argue that current regulations and policies are inadequate to protect against such interdependent privacy violations. The proposed framework relies on the so-called 3Rs: Realize (that some data is being communicated to a third party), Recognize (that the data has privacy implications on others who have rights) and Respect (the rights of others). For the last point (i.e., respect), the authors suggest three solutions: aborting the communication, obtaining consent from others, or modifying (e.g., anonymizing, obfuscating) the data before communicating it. And an interdependent privacy violation occurs when an individual fails one of the 3Rs. The authors further propose four classes of interventions, together with concrete examples, to prevent such failures: “Ensuring realization”, “Encouraging recognition”, “Educating for respect” and “Embracing alternatives”.

5.2 Technical Solutions

In this section, we describe the technical solutions (e.g., architectures, software tools) proposed in the literature to address interdependent privacy issues. Note that there is an active line of research (mostly in the database community) devoted to the case of aggregate queries on correlated records and differential privacy (see Sec. 4.5), including [Chen et al. 2014; Kifer and Machanavajhala 2014; Liu et al. 2016; Song et al. 2017; Yang et al. 2015; Zhu et al. 2015], to name a few. Yet, because these solutions do not focus directly on interdependent privacy, we do not describe them in this survey. We summarize the results of this section in a synthetic table presented at the end of the section (Table 4, on page 31).

Wishart et al. propose an access control solution where data co-owners can arbitrarily restrict the permissions on the co-owned data, i.e., veto voting (see Sec. 3.5) [Wishart et al. 2010]. In case of conflict with a co-owner who applies too many restrictions on the data access, either the restrictive policy is applied as such or the content is modified (i.e., blur the face or crop it) so that this co-owner is no longer affected by the data sharing.

Hu et al. propose a model for detecting and resolving privacy conflicts in OSNs [Hu et al. 2011]. In order to detect potential conflicts for some content, the owner must tag other users involved in the content when publishing the content. The algorithm solves conflicts based on the trust relationships between the users and the sensitivity and visibility of the content; and, finally, it balances data sharing benefits and privacy risks. Finally, the authors implement and evaluate a proof-of-concept prototype in FB and measure its practicality and usability with a user survey of $N = 30$ participants. In their follow-up work, the authors present a multi-party access control model for OSNs [Hu et al. 2013]. In particular, they define four different user roles in their framework: owner (e.g., uploading a content in her own space),

contributor (e.g., uploading the content to someone else's place), stakeholders (e.g., individuals being tagged), and disseminators (e.g., individuals sharing on her own space a content from another space). They aggregate the policies from all the involved users to decide whether to deny or grant access, and they rely on a voting mechanism in the cases of privacy conflicts between users. Finally, they also provide a prototype implementation of their access control model in FB and evaluate it by surveying $N = 35$ participants.

Beato and Peeters propose a collaborative access control based on secret sharing [Beato and Peeters 2014]. The owner encrypts the content with a random key and sends for each co-owner the content, the key (to decrypt the content and take a decision) and a share of the secret. Then co-owners give their share to other users (i.e., the targeted audience) so that only users with a higher number of shares than a defined threshold can decrypt the content.

González-Manzano et al. propose a co-owned personal data management system named CooPeD: it regulates the access control of co-owned objects decomposable into parts, such as photos [González-Manzano et al. 2014] (see Sec. 3.5). With CooPeD, each user involved in the object (e.g., appearing in a photo) can individually manage her privacy preferences and access rights. The proposed approach is evaluated through various means, including a feasibility analysis, a prototype, and a survey of $N = 206$ participants.

Guo et al. propose a framework for collaborative privacy management on *mobile devices* to protect so-called distributed information (e.g., phone numbers appearing on multiple contact lists) [Guo et al. 2014]. Each user can specify which app or app category can access what type of sensitive data. Every user's policy is then enforced on all devices where the user's data appears, assuming all mobile device owners act collaboratively. To test their framework, the authors implement a proof-of-concept prototype on Android and show their practicality with two case studies on contact lists and photo publishing.

Ilia et al. developed a fine-grained tool to regulate the access control of faces in photos [Ilia et al. 2015]. The authors use a three-dimensional access control matrix that defines the object group (i.e., a photo), the objects (i.e., faces), and the subjects (i.e., users). Their approach handles conflicts of privacy policies by modifying the access control's granularity, from the whole photo to the faces of individuals in the photo. Moreover, it automatically identifies users in the photo using face recognition and blurs the faces of the users who have restricted access to it. The authors build a prototype third-party FB app for evaluating its effectiveness. Finally, they interview $N = 52$ users on their perception of existing mechanisms and the newly proposed one. They found out that 77% of them are positive about the proposed approach.

Aditya et al. develop I-Pic, a mobile application for collecting consent from bystanders in photos [Aditya et al. 2016]. It is based on three main principles from a user study: (i) as privacy concerns vary between individuals, privacy policies should be individual too, (ii) privacy policies should be situational, and (iii) users should comply, in general, with the privacy preferences of others. The privacy preferences of bystanders are broadcast via short-range communication (e.g., Bluetooth). The system relies on support vector machine for face recognition of the bystanders and makes use of homomorphic encryption and garbled circuits to protect bystanders' privacy. The authors evaluate the performance of I-Pic, such as accuracy of face detection or runtime, by implementing a prototype on Android.

Mehregan and Fong developed two design patterns for simplifying existing access right schemes for content involving multiple users (referred to as stakeholders) [Mehregan and Fong 2014]. The authors propose different design patterns for scenarios with simple annotations (such as liking, tagging, and sharing) and for those with higher-order annotations (i.e., annotations of annotations, such as replies to comments). Design principles include, among others, that every stakeholder of a content should have a say about the access control policy of this content. The authors evaluate the performance of their approach with real data from LiveJournal and show that it meets the responsiveness requirements of web applications. In a follow-up work, the authors develop a DAC model for shared resources with multiple ownership

that satisfies three criteria of availability: policy satisfiability, feasibility, and resiliency [Mehregan and Fong 2016]. In particular, they extend the ReBAC model (see Sec. 3.5) to reach a mutual access control policy among co-owners through an interactive policy negotiation protocol. The authors also design and evaluate two algorithms that both rely on a SAT solver for verifying policy satisfiability; they further demonstrate their effectiveness for mid-sized organizations.

Such and Rovatsos present an automated method for detecting privacy policy conflicts in OSNs, and they resolve them using negotiation between stakeholders involved in the shared content [Such and Rovatsos 2016]. The proposed approach makes use of ReBAC and of the one-step negotiation protocol. The authors further show that the space of possible compromises grows exponentially with the number of conflicts to be negotiated, and they propose three heuristics to overcome this problem. They evaluate the performance of their different approaches and show that greedy heuristics provide good trade-offs between complexity and optimality when the number of agents increases.

Keküllüoğlu et al. improve upon Such and Rovatsos's work by considering multiple interactions between OSN users. They propose three negotiation strategies including one based on reciprocity between users and that takes into account previous interactions [Keküllüoğlu et al. 2016]. The authors compare their approach with state-of-the-art strategies and, with three synthetic examples, show that reciprocal privacy outperforms other strategies most of the time.

Palomar et al. propose to rely on cryptographic primitives to provide fine-grained access control and co-ownership management in OSNs [Palomar et al. 2016]. In particular, the proposed system relies on anonymous attribute-based credential algorithms and on joint random-secret sharing to ensure confidentiality of users' privacy preferences.



























Rathore and Tripathy propose another multi-party access control model for OSN content [Rathore and Tripathy 2016]. Users are split in groups defined by trust levels assigned by each OSN user to his contacts. When a new content is shared, all the policies of stakeholders (i.e., content (co)-owners) are collected, and access is granted if the requester satisfies a majority of stakeholder policies (that depend on the trust levels of the requester with respect to each stakeholder).

In a follow-up work of [Such and Rovatsos 2016], Misra and Such propose two solutions, PACMAN and REACT, to recommend sharing policies to OSN users, by relying on machine learning techniques, in the case of co-owned content [Misra and Such 2017a,b]. They use the relationship type and strength, and the shared content as attributes for their recommendation agents (i.e., features). Their results show that the proposed agents provide an average accuracy of more than 90%.

Iliia et al. design a collaborative multiparty access control model that ensures that the privacy preferences of all co-owners of a shared content are considered and that their data is not accessible by the service provider and third parties [Iliia et al. 2017]. The shared content is protected from the service provider with encryption by relying on the (k, n) -threshold secret sharing scheme [Shamir 1979]. In order to access a content, a requester then needs to retrieve enough shares from the co-owners' trusted contacts to eventually retrieve the encryption key. Experimental results show that the proposed model does not significantly affect the time to upload and access the shared content.

Guarnieri et al. address the generic problem of preventing probabilistic inference caused by data dependencies in databases [Guarnieri et al. 2017]. To do so, they propose a novel provably secure mechanism for tractable and a practical useful database inference control. This mechanism assumes that the attacker's belief is known, and it ensures that this belief does not increase for all secrets in the security policies. They formalize the various constraints given by security policies by relying on probabilistic logic programming and by making inference tractable by building poly-tree Bayesian networks. Finally, the authors evaluate the efficiency of their scheme with synthetic belief programs containing 1,000 to 100,000 patients and a policy with 100 secrets, and show that it takes less than 0.3 seconds to check a query's security for a database of 100,000 patients.

Table 4. Summary of technical solutions to interdependent privacy issues.

article	data	adversary	technique
[Wishart et al. 2010]	 photo	users	access control (veto voting)
[Hu et al. 2011]	 photo	users	access control (trust, sensitivity, visibility)
[Hu et al. 2013]	 photo	users	access control (weighted voting)
[Beato and Peeters 2014]	 photo	users	cryptography (secret sharing)
[González-Manzano et al. 2014]	 photo	users	access control (fine-grained)
[Guo et al. 2014]	 generic	mobile apps	access control (fine-grained)
[Ilia et al. 2015]	 photo	users	access control (fine-grained) + obfuscation
[Aditya et al. 2016]	 photo	users	cryptography (homomorphic encryption)
[Mehregan and Fong 2014]	 generic	users	access control (discretionary)
[Mehregan and Fong 2016]	 generic	users	access control (relationship-based)
[Such and Rovatsos 2016]	 photo	users	access control (negotiation/action vector)
[Keküllüoğlu et al. 2016]	 photo	users	access control (negotiation/action vector)
[Palomar et al. 2016]	 photo	users	cryptography (attribute-based credential)
[Rathore and Tripathy 2016]	 generic	users	access control (majority voting)
[Misra and Such 2017a,b]	 photo	users	policy recommendation
[Ilia et al. 2017]	 generic	service provider	cryptography (secret sharing)
[Guarnieri et al. 2017]	 medical data	users	access control (inference control)
[Li et al. 2017a]	 photo	users	obfuscation
[Xu et al. 2017]	 photo	users	access control (veto voting)
[Li et al. 2017b]	 photo	users	access control (fine-grained)
[Harkous and Aberer 2017]	 file	cloud storage apps	privacy meter (for users)
[Zhong et al. 2018]	 photo	users	access control (neural network detection)
[Olteanu et al. 2018]	 photo (ext.   )	users	cryptography + face recognition + obfuscation

Li et al. study eight different obfuscation methods for hiding a user in a photo shared on an OSN [Li et al. 2017a]. More specifically, they analyze the trade-off between (i) the effectiveness of the different obfuscation mechanisms, and (ii) the remaining photo utility, i.e., the user satisfaction and experience when looking at the obfuscated photo. The authors compare the different methods by performing a user study with $N = 271$ MTurk users. Their experimental results show that, despite their popularity, blurring and pixelating are ineffective at obfuscating users, and that inpainting (i.e., removing the user entirely and replacing her with something visually consistent with the image) and avatar (i.e., replacing the user with an avatar that preserves some of the elements of the initial representation) are the best options.

Xu et al. propose a privacy-preserving face recognition system for automatically detecting whether someone appears in a photo shared on OSNs [Xu et al. 2017]. In order to protect the privacy of the OSN users, the authors develop a method to locally train their face recognition algorithm. Specifically, they rely on a support vector machine model and train it by discriminating between the users and their graph neighbors (e.g., friends) who are trusted more than random users in the OSN. The locally trained models are then shared among users to obtain a global knowledge and improve each local model in the next iteration. The authors implement a proof-of-concept Android application for FB and show through experimental and theoretical analyses that their distributed approach is both efficient and effective at recognizing OSN user faces.

Li et al. propose a system for preserving individuals' privacy depicted in photos that are shared in instant messaging [Li et al. 2017b]. The faces of stakeholders (i.e., individuals appearing in the photo) are encrypted before the photo is sent and are visible, depending on the policy of each individual with respect to each potential viewer. If the viewer is permitted to see the face of a given stakeholder, they will receive the key of this stakeholder to decrypt the corresponding face. Furthermore, in order to improve user experience, the authors propose to automate the access control process by

exploiting similarities between photos by relying on metrics such as cosine similarity between local features. Finally, the authors show with a proof-of-concept implementation that the performance overhead is limited.

Harkous and Aberer study the issue of interdependent privacy in the context of file sharing with third-party cloud applications [Harkous and Aberer 2017]. They first evaluate the extent of the threat by using a real-world dataset from Google Drive and show that about 37% of the documents are shared with at least one other user. Given these results, the authors propose new privacy indicators that display the actual privacy risks (i.e., what the app has access to already) by taking into account previous permission decisions (of theirs and of their collaborators). The authors assess the effect of their solution with a web experiment with $N = 141$ participants and show that it significantly improves the privacy situation by encouraging users to select the option that minimizes privacy loss.

Zhong et al. automate multiparty privacy conflict detection based on their prior work: a solution to detect privacy levels of a part of a photo [Zhong et al. 2018]. Their solution that relies on a convolutional neural network architecture, obtains good results compared to baseline algorithms (e.g., logistic regression), with 70% accuracy on average.

Olteanu et al. propose ConsenShare, a privacy-preserving system for consensual online sharing of data that have privacy implications for multiple individuals (e.g., photos, DNA), namely interdependent and multiple-subject data [Olteanu et al. 2018]. The work focuses mostly on multimedia data, and more specifically on photos. With ConsenShare, the individuals who appear on a photo are masked until they give consent. Neither the sharing platform operators nor the users can access the masked data. The proposed solution relies on a two-tier architecture (identity and content management services) and cryptographic (i.e., asymmetric encryption) and computer vision techniques (i.e., face recognition). The technical proposal is complemented by a user study to assess the need for and the desirability of the proposed system.

6 CONCLUSIONS AND FUTURE DIRECTIONS

In this survey, we presented a comprehensive review of the interdependent privacy literature, focusing on risks and solutions. This work shows that interdependent privacy, i.e., the situation where the actions of some individuals compromise the privacy of other individuals, has been studied by different communities (including information security and privacy, data science, human-computer interaction/computer-supported cooperative work, and information systems) and mostly in isolation. One of the reasons for this is that these different communities refer to the same situations using different terminologies, such as interdependent privacy, multi-party privacy, and networked privacy.

Regarding the risks, this work shows that a broad range of data types are subject to interdependent privacy risks, including online social network data (i.e., profiles), multimedia data, location data, and genomic data. In order to quantify these risks, most works rely on inference techniques from the machine learning community (e.g., Bayesian networks, support vector machine). Often times, the proposed inference algorithms are ad-hoc and data specific. Most works on interdependent privacy risks are conducted in the information security and privacy or data science communities.

As for the solutions, two broad categories exist: social (cooperative or not) and technical. Most works on cooperative solutions rely on social theories such as the communication privacy management theory and are conducted in the human-computer interaction/computer-supported cooperative work community. These works identify the different coping mechanisms individuals employ to solve online privacy conflicts; they also highlight a lack of tools to assist users to do so on existing online social platforms. Works on non-cooperative solutions rely on game theory and are conducted in the information security and privacy community. Not all such works consider more than two players and only a few extract the parameters of the payoff functions of the game-theoretical models from user data. As for technical solutions, most approaches rely on access control and/or cryptographic techniques, possibly in combination

with data-specific techniques such as face recognition and obfuscation for photos. In fact, many of these works focus on photos. Only a few works consider the service provider as a potential adversary, and they focus only on protecting individuals' privacy with respect to others. Overall, none of the existing solutions is generic enough to handle various data types and adversaries.

We argue that privacy in general, and interdependent privacy in particular, is a highly multi-disciplinary topic and research in this field should be as conducted hand-in-hand between different disciplines. By putting together the works conducted in different communities and by unifying the terminology, this survey enables such a holistic approach. We further learn from our literature review that research on risks and solutions should be addressed in a more principled and less data-dependent manner. We foresee potential promising directions of research towards more holistic frameworks for interdependent privacy research. The general idea is to integrate risk evaluation techniques, technical solutions and human/social aspects more tightly. First, we believe that risk quantification techniques (based on generic graphical models instead of data-specific models) should be better embedded into technical/social solutions in order to assist users/tools to make informed/optimal decisions, by providing them with accurate estimates of the risks involved. Second, we believe that the proposed solutions should account for the interplay between the involved individuals (including service providers as adversaries), both from emotional and rational perspectives, by relying on communication privacy management theory and game theory. Third, we believe that the temporal dimension should be taken into account, especially for medium- and long-term privacy implications of individual actions.

Finally, we argue that interdependent privacy should also be tackled from a legal perspective; although it was mentioned (without using a specific terminology) in Opinion 5/2009 on online social networking²⁹ produced by the Data Protection Working Party,³⁰ it was not explicitly mentioned or addressed in the recently adopted General Data Protection Regulation of the European Union (EU GDPR).

ACKNOWLEDGMENTS

The authors express their sincere gratitude to the anonymous reviewers as well as to Alexandra Mihaela Olteanu and Gergely Biczók for their insightful feedback. The authors also warmly thank Holly Cogliati for her great editing job on the manuscript. This work was partially funded by the Swiss National Science Foundation with Grant #200021_178978.

REFERENCES

2013. Facebook: Where Your Friends Are Your Worst Enemies. <https://packetstormsecurity.com/news/view/22713/Facebook-Where-Your-Friends-Are-Your-Worst-Enemies.html>. (2013). Last visited: Jun. 2019.
- Abigail Abrams. 2018. How an Online DNA Service Revealed the Suspected Golden State Killer. *Time* (2018).
- Alessandro Acquisti and Christina M. Fong. 2012. An Experiment in Hiring Discrimination Via Online Social Networks. *SSRN Electronic Journal* (2012). <https://doi.org/10.2139/ssrn.2031979>
- Paarijaat Aditya, Rijurekha Sen, Peter Druschel, Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele, Bobby Bhattacharjee, and Tong Tong Wu. 2016. I-Pic: A Platform for Privacy-Compliant Image Capture. In *Proc. of MobiSys*. <https://doi.org/10.1145/2906388.2906412>
- Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. 2012. Risks of Friendships on Social Networks. In *Proc. of ICDM*. <https://doi.org/10.1109/ICDM.2012.57>
- Yaqoub Alsarkal, Nan Zhang, and Heng Xu. 2018. Your Privacy Is Your Friend's Privacy: Examining Interdependent Information Disclosure on Online Social Networks. In *Proc. of HICSS*.
- Hadar Aviram. 2012. What Would You Do? Conducting Web-Based Factorial Vignette Surveys. In *Handbook of Survey Methodology for the Social Sciences*, Lior Gideon (Ed.). https://doi.org/10.1007/978-1-4614-3876-2_26
- Erman Ayday and Mathias Humbert. 2017. Inference Attacks against Kin Genomic Privacy. *IEEE Security & Privacy* 15, 5 (2017). <https://doi.org/10.1109/MSP.2017.3681052>

²⁹https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2009/wp163_en.pdf

³⁰an advisory board set up by the European Union for the reform of its data protection laws.

- Michael Backes, Pascal Berrang, Mathias Humbert, Xiaoyu Shen, and Verena Wolf. 2018. Simulating the Large-Scale Erosion of Genomic Privacy Over Time. *IEEE/ACM Trans. on Computational Biology and Bioinformatics* (2018). <https://doi.org/10.1109/TCBB.2018.2859380>
- Michael Backes, Mathias Humbert, Jun Pang, and Yang Zhang. 2017. Walk2friends: Inferring Social Links from Mobility Profiles. In *Proc. of CCS*. <https://doi.org/10.1145/3133956.3133972>
- Filipe Beato and Roel Peeters. 2014. Collaborative Joint Content Sharing for Online Social Networks. In *Proc. of PerCom Workshops*. <https://doi.org/10.1109/PerComW.2014.6815277>
- Justin Lee Becker and Hao Chen. 2009. Measuring Privacy Risk in Online Social Networks. In *Proc. of the Web 2.0 Security and Privacy Workshop (W2SP)*.
- P. Berrang, M. Humbert, Y. Zhang, I. Lehmann, R. Eils, and M. Backes. 2018. Dissecting Privacy Risks in Biomedical Data. In *Proc. of EuroS&P*. <https://doi.org/10.1109/EuroSP.2018.00013>
- Andrew Besmer and Heather Richter Lipford. 2010. Moving beyond Untagging: Photo Privacy in a Tagged World. In *Proc. of CHI*. <https://doi.org/10.1145/1753326.1753560>
- Gergely Biczók and Pern Hui Chia. 2013. Interdependent Privacy: Let Me Share Your Data. In *Proc. of FC*. https://doi.org/10.1007/978-3-642-39884-1_29
- Igor Bilogrevic, Kévin Huguenin, Murtuza Jadliwala, Florent Lopez, Jean-Pierre Hubaux, Philip Ginzboorg, and Valtteri Niemi. 2013. Inferring Social Ties in Academic Networks Using Short-Range Wireless Communications. In *Proc. of WPES*. <https://doi.org/10.1145/2517840.2517842>
- Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast Unfolding of Communities in Large Networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008). <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Edward J. Bloustein. 1978. *Individual and Group Privacy*.
- Violet Blue. 2013a. Anger Mounts after Facebook's 'shadow Profiles' Leak in Bug. <https://www.zdnet.com/article/anger-mounts-after-facebooks-shadow-profiles-leak-in-bug/>. (2013).
- Violet Blue. 2013b. Firm: Facebook's Shadow Profiles Are 'frightening' Dossiers on Everyone. <https://www.zdnet.com/article/firm-facebooks-shadow-profiles-are-frightening-dossiers-on-everyone/>. (2013).
- Danah Boyd. 2012. Networked Privacy. *Surveillance & Society* 10, 3/4 (2012). <https://doi.org/10.24908/ss.v10i3/4.4529>
- Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach. *The Guardian* (2018).
- Barbara Carminati and Elena Ferrari. 2011. Collaborative Access Control in On-Line Social Networks. In *Proc. of CollaborateCom*. <https://doi.org/10.4108/icst.collaboratecom.2011.247109>
- Jin Chen, Jerry Wenjie Ping, Yunjie Calvin Xu, and Bernard C. Y. Tan. 2015. Information Privacy Concern About Peer Disclosure in Online Social Networks. *IEEE Trans. on Engineering Management* 62, 3 (2015). <https://doi.org/10.1109/TEM.2015.2432117>
- Rui Chen, Benjamin C. M. Fung, Philip S. Yu, and Bipin C. Desai. 2014. Correlated Network Data Publication via Differential Privacy. *The VLDB Journal* 23, 4 (2014). <https://doi.org/10.1007/s00778-013-0344-8>
- Hichang Cho and Anna Filippova. 2016. Networked Privacy Management in Facebook: A Mixed-Methods and Multinational Study. In *Proc. of CSCW*. <https://doi.org/10.1145/2818048.2819996>
- Ben C. F. Choi, Zhenhui (Jack) Jiang, Bo Xiao, and Sung S. Kim. 2015. Embarrassing Exposures in Online Social Networks: An Integrated Perspective of Privacy Invasion and Relationship Bonding. *Information Systems Research* 26, 4 (2015). <https://doi.org/10.1287/isre.2015.0602>
- Chun Fung Choi and Zhenhui Jiang. 2013. Trading Friendship for Value: An Investigation of Collective Privacy Concerns in Social Application Usage. In *Proc. of ICIS*.
- Tharntip Tawnie Chutikulrungeee and Oliver Kisalay Burmeister. 2017. Interdependent Privacy. *ORBIT Journal* 1, 2 (2017). <https://doi.org/10.29297/orbit.v1i2.38>
- Edward H Clarke. 1971. Multipart Pricing of Public Goods. *Public choice* 11, 1 (1971).
- D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. Inferring Social Ties from Geographic Coincidences. *Proceedings of the National Academy of Sciences (PNAS)* 107, 52 (2010). <https://doi.org/10.1073/pnas.1006155107>
- Stan Damen, Nicola Zannone, Stan Damen, and Nicola Zannone. 2014. Privacy Implications of Privacy Settings and Tagging in Facebook. In *Proc. of SDM*. https://doi.org/10.1007/978-3-319-06811-4_16
- Emiliano De Cristofaro. 2013. An Exploratory Ethnographic Study of Issues and Concerns with Whole Genome Sequencing. *arXiv:1306.4962 [cs, q-bio]* (2013). <https://doi.org/10.14722/usec.2014.23020> [arXiv:cs, q-bio/1306.4962](https://arxiv.org/abs/1306.4962)
- Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. 2009. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication* 15, 1 (2009). <https://doi.org/10.1111/j.1083-6101.2009.01494.x>
- Ratan Dey, Yuan Ding, and Keith W. Ross. 2013. Profiling High-School Students with Facebook: How Online Privacy Laws Can Actually Increase Minors' Risk. In *Proc. of IMC*. <https://doi.org/10.1145/2504730.2504733>
- Ratan Dey, Cong Tang, Keith Ross, and Nitesh Saxena. 2012. Estimating Age Privacy Leakage in Online Social Networks. In *Proc. of INFOCOM*. <https://doi.org/10.1109/INFOCOM.2012.6195711>
- Iman Deznabi, Mohammad Mobayen, Nazanin Jafari, Ozgur Tastan, and Erman Ayday. 2018. An Inference Attack on Genomic Data Using Kinship, Complex Correlations, and Phenotype Information. *IEEE/ACM Trans. on Computational Biology and Bioinformatics* 15, 4 (2018). <https://doi.org/10.1109/TCBB.2017.2709740>
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Prof. of the Conf. on Theory of Cryptography (TCC)*. https://doi.org/10.1007/11681878_14

- Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. 2009. Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences (PNAS)* 106, 36 (2009). <https://doi.org/10.1073/pnas.0900282106>
- Taher Elgamal. 1985. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. *IEEE Trans. on Information Theory* 31, 4 (1985). <https://doi.org/10.1109/TIT.1985.1057074>
- Luciano Floridi. 2014. Open Data, Data Protection, and Group Privacy. *Philosophy & Technology* 27, 1 (2014). <https://doi.org/10.1007/s13347-014-0157-8>
- Ricard L. Fogues, Pradeep K. Murukannaiah, Jose M. Such, and Munindar P. Singh. 2017a. Sharing Policies in Multiuser Privacy Scenarios: Incorporating Context, Preferences, and Arguments in Decision Making. *ACM Trans. on Computer-Human Interaction* 24, 1 (2017). <https://doi.org/10.1145/3038920>
- Ricard L. Fogues, Pradeep K. Murukannaiah, Jose M. Such, and Munindar P. Singh. 2017b. SoSharP: Recommending Sharing Policies in Multiuser Privacy Scenarios. *IEEE Internet Computing* 21, 6 (2017). <https://doi.org/10.1109/MIC.2017.4180836>
- Ricard L. Fogues, Pradeep Murukannaiah, Jose M. Such, Agustin Espinosa, Ana Garcia-Fornes, and Munindar Singh. 2015. Argumentation for Multi-Party Privacy Management. In *Proc. of ACySe*.
- Philip WL Fong. 2011. Relationship-Based Access Control: Protection Model and Policy Language. In *Proc. of CODASPY*. <https://doi.org/10.1145/1943513.1943539>
- Siyao Fu, Haibo He, and Zeng-Guang Hou. 2014. Learning Race from Face: A Survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 36, 12 (2014). <https://doi.org/10.1109/TPAMI.2014.2321570>
- David Gefen, Detmar Straub, and Marie-Claude Boudreau. 2000. Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the association for information systems* 4, 1 (2000). <https://doi.org/10.17705/1CAIS.00407>
- Stefania Gnesi, Ilaria Matteucci, Corrado Moiso, Paolo Mori, Marinella Petrocchi, and Michele Vescovi. 2014. My Data, Your Data, Our Data: Managing Privacy Preferences in Multiple Subjects Personal Data. In *Proc. of APF*, Vol. 8450. https://doi.org/10.1007/978-3-319-06749-0_11
- Neil Zhenqiang Gong and Bin Liu. 2016. You Are Who You Know and How You Behave: Attribute Inference Attacks via Users Social Friends and Behaviors. In *Proc. of USENIX Security*.
- Neil Zhenqiang Gong, Ameet Talwalkar, Lester Mackey, Ling Huang, Eui Chul Richard Shin, Emil Stefanov, Elaine (Runting) Shi, and Dawn Song. 2014. Joint Link Prediction and Attribute Inference Using a Social-Attribute Network. *ACM Trans. on Intelligent Systems and Technology* 5, 2 (2014). <https://doi.org/10.1145/2594455>
- Lorena González-Manzano, Ana I. González-Tablas, José M. de Fuentes, and Arturo Ribagorda. 2014. CooPeD: Co-Owned Personal Data Management. *Computers & Security* 47 (2014). <https://doi.org/10.1016/j.cose.2014.06.003>
- Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. 2006. Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data. In *Proc. of CCS*. <https://doi.org/10.1145/1180405.1180418>
- Marco Guarnieri, Srdjan Marinovic, and David Basin. 2017. Securing Databases from Probabilistic Inference. In *Proc. of CSF*. <https://doi.org/10.1109/CSF.2017.30>
- Pritam Gundecha, Geoffrey Barbier, and Huan Liu. 2011. Exploiting Vulnerability to Secure User Privacy on a Social Networking Site. In *Proc. of KDD*. <https://doi.org/10.1145/2020408.2020489>
- Pritam Gundecha, Geoffrey Barbier, Jiliang Tang, and Huan Liu. 2014. User Vulnerability and Its Reduction on a Social Networking Site. *ACM Trans. on Knowledge Discovery from Data* 9, 2 (2014). <https://doi.org/10.1145/2630421>
- Yao Guo, Li Zhang, and Xiangqun Chen. 2014. Collaborative Privacy Management: Mobile Privacy beyond Your Own Devices. In *Proc. of SPME*. <https://doi.org/10.1145/2646584.2646590>
- Hamza Harkous and Karl Aberer. 2017. "If You Can't Beat Them, Join Them": A Usability Approach to Interdependent Privacy in Cloud Apps. In *Proc. of CODASPY*. <https://doi.org/10.1145/3029806.3029837>
- Jianming He, Wesley W. Chu, and Zhenyu Liu. 2006. Inferring Privacy Information from Social Networks. In *Proc. of ISI*. https://doi.org/10.1007/11760146_14
- Benjamin Henne, Christian Szongott, and Matthew Smith. 2013. SnapMe If You Can: Privacy Threats of Other Peoples' Geo-Tagged Media and What We Can Do about It. In *Proc. of WiSec*. <https://doi.org/10.1145/2462096.2462113>
- Benjamin Mako Hill. 2014. Google Has Most of My Email Because It Has All of Yours. <https://mako.cc/copyrighteous/google-has-most-of-my-email-because-it-has-all-of-yours>. (2014). Last visited: Feb. 2019.
- Hongxin Hu and Gail-Joon Ahn. 2011. Multiparty Authorization Framework for Data Sharing in Online Social Networks. In *Proc. of DBSec (Lecture Notes in Computer Science)*. https://doi.org/10.1007/978-3-642-22348-8_5
- Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2011. Detecting and Resolving Privacy Conflicts for Collaborative Data Sharing in Online Social Networks. In *Proc. of ACSAC*. <https://doi.org/10.1145/2076732.2076747>
- Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2013. Multiparty Access Control for Online Social Networks: Model and Mechanisms. *IEEE Trans. on Knowledge and Data Engineering* 25, 7 (2013). <https://doi.org/10.1109/TKDE.2012.97>
- Hongxin Hu, Gail-Joon Ahn, Ziming Zhao, and Dejun Yang. 2014. Game Theoretic Analysis of Multiparty Access Control in Online Social Networks. In *Proc. of SACMAT*. <https://doi.org/10.1145/2613087.2613097>
- Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. 2013a. Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy. In *Proc. of CCS*. <https://doi.org/10.1145/2508859.2516707>
- Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. 2015. On Non-Cooperative Genomic Privacy. In *Proc. of FC*. https://doi.org/10.1007/978-3-662-47854-7_24

- Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. 2017. Quantifying Interdependent Risks in Genomic Privacy. *ACM Trans. on Privacy and Security* 20, 1 (2017). <https://doi.org/10.1145/3035538>
- Mathias Humbert, Didier Dupertuis, and Kévin Huguenin. 2019. *Data-Less Evaluation of Kin Genomic Privacy*. Technical Report.
- Mathias Humbert, Théophile Studer, Matthias Grossglauser, and Jean-Pierre Hubaux. 2013b. Nowhere to Hide: Navigating around Privacy in Online Social Networks. In *Proc. of ESORICS*. https://doi.org/10.1007/978-3-642-40203-6_38
- Panagiotis Ilija, Barbara Carminati, Elena Ferrari, Paraskevi Fragopoulou, and Sotiris Ioannidis. 2017. SAMPAC: Socially-Aware Collaborative Multi-Party Access Control. In *Proc. of CODASPY*. <https://doi.org/10.1145/3029806.3029834>
- Panagiotis Ilija, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proc. of CCS*. <https://doi.org/10.1145/2810103.2813603>
- Carter Jernigan and Behram F. T. Mistree. 2009. Gaydar: Facebook Friendships Expose Sexual Orientation. *First Monday* 14, 10 (2009). <https://doi.org/10.5210/fin.v14i10.2611>
- Haiyan Jia and Heng Xu. 2016a. Autonomous and Interdependent: Collaborative Privacy Management on Social Networking Sites. In *Proc. of CHI*. <https://doi.org/10.1145/2858036.2858415>
- Haiyan Jia and Heng Xu. 2016b. Measuring Individuals' Concerns over Collective Privacy on Social Networking Sites. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10, 1 (2016). <https://doi.org/10.5817/CP2016-1-4>
- Jinyuan Jia, Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. 2017. AttrInfer: Inferring User Attributes in Online Social Networks Using Markov Random Fields. In *Proc. of WWW*. <https://doi.org/10.1145/3038912.3052695>
- David Jurgens. 2013. That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships. In *Proc. of ICWSM*.
- Gulce Kale, Erman Ayday, and Ozgur Tastan. 2018. A Utility Maximizing and Privacy Preserving Approach for Protecting Kinship in Genomic Databases. *Bioinformatics* 34, 2 (2018). <https://doi.org/10.1093/bioinformatics/btx568>
- Bernadette Kamleitner and Vince Mitchell. 2019. Your Data Is My Data: A Framework for Addressing Interdependent Privacy Infringements. *Journal of Public Policy & Marketing* (2019). <https://doi.org/10.1177/0743915619858924>
- Dilara Keküllüoğlu, Nadin Kökciyan, and Pinar Yolum. 2016. Strategies for Privacy Negotiation in Online Social Networks. In *Proc. of PrAISE*. <https://doi.org/10.1145/2970030.2970035>
- Daniel Kifer and Ashwin Machanavajhala. 2011. No Free Lunch in Data Privacy. In *Proc. of SIGMOD*. <https://doi.org/10.1145/1989323.1989345>
- Daniel Kifer and Ashwin Machanavajhala. 2014. Pufferfish: A Framework for Mathematical Privacy Definitions. *ACM Trans. on Database Systems* 39, 1 (2014). <https://doi.org/10.1145/2514689>
- William S Klug, Michael R Cummings, Charlotte Spencer, and Sarah M Ward. 2003. *Concepts of Genetics*. OCLC: 909434618.
- Daphne Koller and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*.
- Gergely Kotyuk and Levente Buttyan. 2012. A Machine Learning Based Approach for Predicting Undisclosed Attributes in Social Networks. In *Proc. of PerCom Workshops*. <https://doi.org/10.1109/PerComW.2012.6197511>
- Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proc. of CHI*. <https://doi.org/10.1145/1978942.1979420>
- Aron Laszka, Mark Felegyhazi, and Levente Buttyan. 2014. A Survey of Interdependent Information Security Games. *Comput. Surveys* 47, 2 (2014). <https://doi.org/10.1145/2635673>
- Gil Levi and Tal Hassner. 2015. Age and Gender Classification Using Convolutional Neural Networks. In *Proc. of CVPRW*. <https://doi.org/10.1109/CVPRW.2015.7301352>
- Fenghua Li, Jingyang Yu, Lingcui Zhang, Zhe Sun, and Mengfan Lv. 2017b. A Privacy-Preserving Method for Photo Sharing in Instant Message Systems. In *Proc. of ICCSP*. <https://doi.org/10.1145/3058060.3058081>
- Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017a. Effectiveness and Users' Experience of Obfuscation as a Privacy-Enhancing Technology for Sharing Photos. In *Proc. of CSCW*, Vol. 1. <https://doi.org/10.1145/3134702>
- Kaitai Liang, Joseph K. Liu, Rongxing Lu, and Duncan S. Wong. 2015. Privacy Concerns for Photo Sharing in Online Social Networks. *IEEE Internet Computing* 19, 2 (2015). <https://doi.org/10.1109/MIC.2014.107>
- Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. 2009. Inferring Private Information Using Social Network Data. In *Proc. of WWW*. <https://doi.org/10.1145/1526709.1526899>
- Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. 2016. Dependence Makes You Vulnerable: Differential Privacy Under Dependent Tuples. In *Proc. of NDSS*. <https://doi.org/10.14722/ndss.2016.23279>
- A. Machanavajhala, J. Gehrke, D. Kifer, and M. Venkatasubramanian. 2006. L-Diversity: Privacy beyond k-Anonymity. In *Proc. of ICDE*. <https://doi.org/10.1109/ICDE.2006.1>
- Ani Manichaikul, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michèle Sale, and Wei-Min Chen. 2010. Robust Relationship Inference in Genome-Wide Association Studies. *Bioinformatics* 26, 22 (2010). <https://doi.org/10.1093/bioinformatics/btq559>
- Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Başar, and Jean-Pierre Hubaux. 2013. Game Theory Meets Network Security and Privacy. *Comput. Surveys* 45, 3 (2013). <https://doi.org/10.1145/2480741.2480742>
- Alice E Marwick and Danah Boyd. 2014. Networked Privacy: How Teenagers Negotiate Context in Social Media. *New Media & Society* 16, 7 (2014). <https://doi.org/10.1177/1461444814543995>

- Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (2001). <https://doi.org/10.1146/annurev.soc.27.1.415>
- Pooya Mehregan and Philip W.L. Fong. 2016. Policy Negotiation for Co-Owned Resources in Relationship-Based Access Control. In *Proc. of SACMAT*. <https://doi.org/10.1145/2914642.2914652>
- Pooya Mehregan and Philip W. L. Fong. 2014. Design Patterns for Multiple Stakeholders in Social Computing. In *Proc. of DBSec*, Vol. 8566. https://doi.org/10.1007/978-3-662-43936-4_11
- Dominic Meier, Yvonne Anne Pignolet, Stefan Schmid, and Roger Wattenhofer. 2014. On the Windfall and Price of Friendship: Inoculation Strategies on Social Networks. *Computer Networks* 62 (2014). <https://doi.org/10.1016/j.bjp.2013.12.004>
- Alfred J. Menezes, Paul C. Van Oorschot, and Scott A. Vanstone. 1997. *Handbook of Applied Cryptography*.
- Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. 2010. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *Proc. of WSDM*. <https://doi.org/10.1145/1718487.1718519>
- Gaurav Misra and Jose M. Such. 2017a. PACMAN: Personal Agent for Access Control in Social Media. *IEEE Internet Computing* 21, 6 (2017). <https://doi.org/10.1109/MIC.2017.4180831>
- Gaurav Misra and Jose M. Such. 2017b. REACT: REcommending Access Control Decisions To Social Media Users. In *Proc. of ASONAM*. <https://doi.org/10.1145/3110025.3110073>
- Mainack Mondal, Peter Druschel, Krishna P. Gummadi, and Alan Mislove. 2014. Beyond Access Control: Managing Online Privacy via Exposure. In *Proc. of USEC*. <https://doi.org/10.14722/usec.2014.23046>
- Heather Murphy. 2018. Genealogists Turn to Cousins' DNA and Family Trees to Crack Five More Cold Cases. *The New York Times* (2018).
- Roger B. Myerson. 2004. *Game Theory: Analysis of Conflict* (6. print ed.).
- Arvind Narayanan and Vitaly Shmatikov. 2005. Obfuscated Databases and Group Privacy. In *Proc. of CCS*. <https://doi.org/10.1145/1102120.1102135>
- Helen Fay Nissenbaum. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*.
- Alexandra-Mihaela Olteanu, Kévin Huguenin, Italo Dacosta, and Jean-Pierre Hubaux. 2018. Consensual and Privacy-Preserving Sharing of Multi-Subject and Interdependent Data. In *Proc. of NDSS*. <https://doi.org/10.14722/ndss.2018.23002>
- Alexandra-Mihaela Olteanu, Kévin Huguenin, Reza Shokri, and Jean-Pierre Hubaux. 2014. Quantifying the Effect of Co-Locations on Location Privacy. In *Proc. of PETS*. https://doi.org/10.1007/978-3-319-08506-7_10
- Alexandra-Mihaela Olteanu, Kévin Huguenin, Reza Shokri, Mathias Humbert, and Jean-Pierre Hubaux. 2017. Quantifying Interdependent Privacy Risks with Location Data. *IEEE Trans. on Mobile Computing* 16, 3 (2017). <https://doi.org/10.1109/TMC.2016.2561281>
- Alexandra-Mihaela Olteanu, Mathias Humbert, Kévin Huguenin, and Jean-Pierre Hubaux. 2019. The (Co)-Location Sharing Game. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2019, 2 (2019). <https://doi.org/10.2478/popets-2019-0017>
- Zafer D. Ozdemir, H. Jeff Smith, and John H. Benamati. 2017. Antecedents and Outcomes of Information Privacy Concerns in a Peer Context: An Exploratory Study. *European Journal of Information Systems* 26, 6 (2017). <https://doi.org/10.1057/s41303-017-0056-z>
- Federica Paci, Anna Squicciarini, and Nicola Zannone. 2018. Survey on Access Control for Community-Centered Collaborative Systems. *Comput. Surveys* 51, 1 (2018). <https://doi.org/10.1145/3146025>
- Pascal Paillier. 1999. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In *Proc. of EUROCRYPT*, Vol. 1592. https://doi.org/10.1007/3-540-48910-X_16
- Esther Palomar, Álvaro Galán, Almudena Alcaide, and Lorena González-Manzano. 2016. Implementing a Privacy-Enhanced Attribute-Based Credential System for Online Social Networks with Co-Ownership Management. *IET Information Security* 10, 2 (2016). <https://doi.org/10.1049/iet-ifs.2014.0466>
- Judea Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.
- João Paulo Pesce, Diego Las Casas, Gustavo Rauber, and Virgílio Almeida. 2012. Privacy Attacks in Social Media Using Photo Tagging Networks: A Case Study with Facebook. In *Proc. of PSOSM*. <https://doi.org/10.1145/2185354.2185358>
- Sandra Petronio. 2010. Communication Privacy Management Theory: What Do We Know about Family Privacy Regulation? *Journal of Family Theory & Review* 2, 3 (2010). <https://doi.org/10.1111/j.1756-2589.2010.00052.x>
- Sandra Sporbert Petronio. 2002. *Boundaries of Privacy: Dialectics of Disclosure*.
- Yu Pu and Jens Grossklags. 2014. An Economic Model and Simulation Results of App Adoption Decisions on Networks with Interdependent Privacy Consequences. In *Proc. of GameSec*. https://doi.org/10.1007/978-3-319-12601-2_14
- Yu Pu and Jens Grossklags. 2015. Using Conjoint Analysis to Investigate the Value of Interdependent Privacy in Social App Adoption Scenarios. In *Proc. of ICIS*.
- Yu Pu and Jens Grossklags. 2016. Towards a Model on the Factors Influencing Social App Users' Valuation of Interdependent Privacy. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2016, 2 (2016). <https://doi.org/10.1515/popets-2016-0005>
- Yu Pu and Jens Grossklags. 2017. Valuating Friends' Privacy: Does Anonymity of Sharing Personal Data Matter?. In *Proc. of SOUPS*.
- Andrew Quodling. 2018. Shadow Profiles - Facebook Knows about You, Even If You're Not on Facebook. <http://theconversation.com/shadow-profiles-facebook-knows-about-you-even-if-youre-not-on-facebook-94804>. (2018).
- Laura Radaelli, Piotr Sapiezynski, Florimond Houssiau, Erez Shmueli, and Yves-Alexandre de Montjoye. 2018. Quantifying Surveillance in the Networked Age: Node-Based Intrusions and Group Privacy. *arXiv:1803.09007 [cs]* (2018). arXiv:cs/1803.09007
- Nemi Chandra Rathore and Somanath Tripathy. 2016. Collaborative Access Control Model for Online Social Networks. In *Proc. of IACC*. <https://doi.org/10.1109/IACC.2016.14>

- Aruneer Ratikan and Mikifumi Shikida. 2014. Privacy Protection Based Privacy Conflict Detection and Solution in Online Social Networks. In *Proc. of HAS*. https://doi.org/10.1007/978-3-319-07620-1_38
- Kate Raynes-Goldie. 2010. Aliases, Creeping, and Wall Cleaning: Understanding Privacy in the Age of Facebook. *First Monday* 15, 1 (2010).
- Elissa M. Redmiles, Michelle L. Mazurek, and John P. Dickerson. 2018. Dancing Pigs or Externalities?: Measuring the Rationality of Security Decisions. In *Proc. of EC*. <https://doi.org/10.1145/3219166.3219185>
- Eunsu Ryu, Yao Rong, Jie Li, and Ashwin Machanavajjhala. 2013. Curso: Protect Yourself from Curse of Attribute Inference. In *Proc. of DBSocial*. <https://doi.org/10.1145/2484702.2484706>
- Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham. 2012. Finding Your Friends and Following Them to Where You Are. In *Proc. of WSDM*. <https://doi.org/10.1145/2124295.2124380>
- Amit Sahai and Brent Waters. 2005. Fuzzy Identity-Based Encryption. In *Proc. of EUROCRYPT*, Vol. 3494. https://doi.org/10.1007/11426639_27
- Sriram Sankararaman, Guillaume Obozinski, Michael I Jordan, and Eran Halperin. 2009. Genomic Privacy and Limits of Individual Detection in a Pool. *Nature Genetics* 41, 9 (2009). <https://doi.org/10.1038/ng.436>
- Emre Sarigol, David Garcia, and Frank Schweitzer. 2014. Online Privacy as a Collective Phenomenon. In *Proc. of COSN*. <https://doi.org/10.1145/2660460.2660470>
- Adi Shamir. 1979. How to Share a Secret. *Commun. ACM* 22, 11 (1979).
- Yan Shoshitaishvili, Christopher Kruegel, and Giovanni Vigna. 2015. Portrait of a Privacy Invasion. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2015, 1 (2015). <https://doi.org/10.1515/popets-2015-0004>
- Suyash S. Shringarpure and Carlos D. Bustamante. 2015. Privacy Risks from Genomic Data-Sharing Beacons. *The American Journal of Human Genetics* 97, 5 (2015). <https://doi.org/10.1016/j.ajhg.2015.09.010>
- Shuang Song, Yizhen Wang, and Kamalika Chaudhuri. 2017. Pufferfish Privacy Mechanisms for Correlated Data. In *Proc. of SIGMOD*. <https://doi.org/10.1145/3035918.3064025>
- Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. 2009. Collective Privacy Management in Social Networks. In *Proc. of WWW*. <https://doi.org/10.1145/1526709.1526780>
- Anna C. Squicciarini, Mohamed Shehab, and Joshua Wede. 2010. Privacy Policies for Shared Content in Social Network Sites. *The VLDB Journal* 19, 6 (2010). <https://doi.org/10.1007/s00778-010-0193-7>
- Anna C. Squicciarini, Heng Xu, and Xiaolong Luke Zhang. 2011. CoPE: Enabling Collaborative Privacy Management in Online Social Networks. *Journal of the American Society for Information Science and Technology* (2011). <https://doi.org/10.1002/asi.21473>
- Frank Stajano, Lucia Bianchi, Pietro Liò, and Douwe Korff. 2008. Forensic Genomics: Kin Privacy, Driftnets and Other Open Questions. In *Proc. of WPES*. <https://doi.org/10.1145/1456403.1456407>
- Jose M. Such and Natalia Criado. 2014. Adaptive Conflict Resolution Mechanism for Multi-Party Privacy Management in Social Media. In *Proc. of WPES*. <https://doi.org/10.1145/2665943.2665964>
- Jose M. Such and Natalia Criado. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Trans. on Knowledge and Data Engineering* 28, 7 (2016). <https://doi.org/10.1109/TKDE.2016.2539165>
- Jose M. Such and Natalia Criado. 2018. Multiparty Privacy in Social Media. *Commun. ACM* 61, 8 (2018). <https://doi.org/10.1145/3208039>
- Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-Scale Empirical Study. In *Proc. of CHI*. <https://doi.org/10.1145/3025453.3025668>
- Jose M. Such and Michael Rovatsos. 2016. Privacy Policy Negotiation in Social Media. *ACM Trans. on Autonomous and Adaptive Systems* 11, 1 (2016). <https://doi.org/10.1145/2821512>
- Arun Sundararajan. 2008. Local Network Effects and Complex Network Structure. *The B.E. Journal of Theoretical Economics* 7, 1 (2008). <https://doi.org/10.2202/1935-1704.1319>
- Latanya Sweeney. 2002. K-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002). <https://doi.org/10.1142/S0218488502001648>
- Iraklis Symeonidis, Gergely Biczók, Fatemeh Shirazi, Cristina Pérez-Solà, Jessica Schroers, and Bart Preneel. 2018. Collateral Damage of Facebook Third-Party Applications: A Comprehensive Study. *Computers & Security* 77 (2018). <https://doi.org/10.1016/j.cose.2018.03.015>
- Iraklis Symeonidis, Fatemeh Shirazi, Gergely Biczók, Cristina Pérez-Solà, and Bart Preneel. 2016a. Collateral Damage of Facebook Apps: Friends, Providers, and Privacy Interdependence. In *Proc. of SEC*. https://doi.org/10.1007/978-3-319-33630-5_14
- Iraklis Symeonidis, Pagona Tsormpatzoudi, and Bart Preneel. 2016b. Collateral Damage of Online Social Network Applications. In *Proc. of ICISSP*. <https://doi.org/10.5220/0005806705360541>
- Kurt Thomas, Chris Grier, and David Nicol. 2010. Unfriendly: Multi-Party Privacy Risks in Social Networks. In *Proc. of PETS*. https://doi.org/10.1007/978-3-642-14527-8_14
- Timothy Thornton, Hua Tang, Thomas J. Hoffmann, Heather M. Ochs-Balcom, Bette J. Caan, and Neil Risch. 2012. Estimating Kinship in Admixed Populations. *The American Journal of Human Genetics* 91 (2012). <https://doi.org/10.1016/j.ajhg.2012.05.024>
- Jessica Vitak, Pamela Wisniewski, Xinru Page, Airi Lampinen, Eden Litt, Ralf De Wolf, Patrick Gage Kelley, and Manya Sleeper. 2015. The Future of Networked Privacy: Challenges and Opportunities. In *Proc. of CSCW*. <https://doi.org/10.1145/2685553.2685554>
- John Von Neumann and Oskar Morgenstern. 2007. *Theory of Games and Economic Behavior* (60th anniversary ed.).

- Nevena Vratonjic, Kévin Huguenin, Vincent Bindschaedler, and Jean-Pierre Hubaux. 2013. How Others Compromise Your Location Privacy: The Case of Public Hotspots. In *Proc. of PETS*. https://doi.org/10.1007/978-3-642-39077-7_7
- Nevena Vratonjic, Kévin Huguenin, Vincent Bindschaedler, and Jean-Pierre Hubaux. 2014. A Location-Privacy Threat Stemming from the Use of Shared Public IP Addresses. *IEEE Trans. on Mobile Computing* 13, 11 (2014). <https://doi.org/10.1109/TMC.2014.2309953>
- Isabel Wagner and David Eckhoff. 2018. Technical Privacy Metrics: A Systematic Survey. *Comput. Surveys* 51, 3 (2018). <https://doi.org/10.1145/3168389>
- Na Wang, Heng Xu, and Jens Grossklags. 2011b. Third-Party Apps on Facebook: Privacy and the Illusion of Control. In *Proc. of CHIMIT*. <https://doi.org/10.1145/2076444.2076448>
- Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011a. "I Regretted the Minute I Pressed Share": A Qualitative Study of Regrets on Facebook. In *Proc. of SOUPS*. <https://doi.org/10.1145/2078827.2078841>
- J. Weidman, W. Aurite, and J. Grossklags. 2018. On Sharing Intentions, and Personal and Interdependent Privacy Considerations for Genetic Data: A Vignette Study. *IEEE/ACM Trans. on Computational Biology and Bioinformatics* (2018). <https://doi.org/10.1109/TCBB.2018.2854785>
- Alan F. Westin. 1970. *Privacy and Freedom*.
- Ryan Wishart, Domenico Corapi, Srdjan Marinovic, and Morris Sloman. 2010. Collaborative Privacy Policy Authoring in a Social Networking Context. In *Proc. of POLICY*. <https://doi.org/10.1109/POLICY.2010.13>
- Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for My Space: Coping Mechanisms for SNS Boundary Regulation. In *Proc. of CHI*. <https://doi.org/10.1145/2207676.2207761>
- Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. 2017. My Privacy My Decision: Control of Photo Sharing on Online Social Networks. *IEEE Trans. on Dependable and Secure Computing* 14, 2 (2017). <https://doi.org/10.1109/TDSC.2015.2443795>
- Bin Yang, Issei Sato, and Hiroshi Nakagawa. 2015. Bayesian Differential Privacy on Correlated Data. In *Proc. of SIGMOD*. <https://doi.org/10.1145/2723372.2747643>
- Lingjing Yu, Sri Mounica Motipalli, Dongwon Lee, Peng Liu, Heng Xu, Qingyun Liu, Jianlong Tan, and Bo Luo. 2018. My Friend Leaks My Privacy: Modeling and Analyzing Privacy in Social Networks. In *Proc. of SACMAT*. <https://doi.org/10.1145/3205977.3205981>
- Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. In *Proc. of ICWSM*.
- Elena Zheleva and Lise Getoor. 2009. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *Proc. of WWW*. <https://doi.org/10.1145/1526709.1526781>
- Haoti Zhong, Anna Squicciarini, and David Miller. 2018. Toward Automated Multiparty Privacy Conflict Detection. In *Proc. of CIKM*. <https://doi.org/10.1145/3269206.3269329>
- Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You Are Where You Go: Inferring Demographic Attributes from Location Check-Ins. In *Proc. of WSDM*. <https://doi.org/10.1145/2684822.2685287>
- Tianqing Zhu, Ping Xiong, Gang Li, and Wanlei Zhou. 2015. Correlated Differential Privacy: Hiding Information in Non-IID Data Set. *IEEE Trans. on Information Forensics and Security* 10, 2 (2015). <https://doi.org/10.1109/TIFS.2014.2368363>