Selected Privacy Research Topics in the ARES Project: An Overview

Jesús A. Manjón and Josep Domingo-Ferrer

Universitat Rovira i Virgili
UNESCO Chair in Data Privacy,
Department of Computer Engineering and Mathematics,
Av. Països Catalans 26, E-43007 Tarragona, Catalonia
{jesus.manjon,josep.domingo}@urv.cat

Abstract. This chapter gives an overview of some of the data privacy research carried out by the team at Universitat Rovira i Virgili within the ARES project. Topics reviewed include query profile privacy, location privacy, differential privacy and anti-discrimination.

1 Introduction

Data privacy is the adaptation to the Information Society of the fundamental right to privacy and private life, included by the United Nations in the Universal Declaration of Human Rights (1948), whose Article 12 states: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks".

The rise of information technologies has arisen new threats against personal privacy such as profiling, location tracking or reidentification. Data privacy technologies are about technically enforcing the above right in the information society.

In this chapter we give a general overview of the data privacy research carried out by the team at Universitat Rovira i Virgili. The topics covered here relate to privacy in databases: user privacy (*i.e.* query profile privacy), respondent privacy (*i.e.* data anonymization) and anti-discrimination protection in data mining.

Section 2 deals with query profile privacy, Section 3 covers location privacy in location-based services. Work to increase the utility of data sets anonymized under the differential privacy model is reported in Section 4. Research on methods to protect against discrimination in data mining are covered in 5. Finally, conclusions are drawn in Section 6.

2 Query profile privacy

Data bases and web search engines (WSE, *e.g.* Google, Bing, etc.) are widely used to find a certain piece of data among a huge amount of information in the shortest possible time. However, these useful tools also pose a privacy threat to the users:

database servers and WSE may profile their users by storing and analyzing past searches submitted by them. To address this privacy threat, several solutions have been proposed in the literature.

Private information retrieval (PIR) is possibly the most ambitious solution, but it falls short of being practically deployable. In PIR, a user wants to retrieve an item from a database or search engine without the latter learning which item the user is interested in. PIR was invented in 1995 by Chor *et al.* [5, 7] with the assumption that there are at least two copies of the same database, which do not communicate with each other. In the PIR literature, the database is usually modeled as a vector and it is assumed that the user knows the physical address of the sought item. Keyword PIR [6] is a more flexible form of PIR: the user can submit a query consisting of a keyword and no modification in the structure of the database is needed.

However, if one wishes to run PIR against a search engine, there are some fundamental shortcomings: (i) the server has no motivation to co-operate in the PIR protocol; (ii) it is not realistic to model a database (let alone the world-wide web) as a vector in which the user can be assumed to know the physical location of the keyword sought. Even keyword PIR does not really fit, as it still assumes a mapping between individual keywords and physical addresses (in fact, each keyword is used as an alias of a physical address). A WSE allowing only searches of individual keywords stored in this way would be much more limited than real engines like Google or Yahoo.

In the sequel, we give solutions that relax the property of PIR that the database or WSE should not know the item retrieved by the user. We restrict our ambitions to allowing the user to keep her query profile (query history) confidential. We start with standalone solutions, in which a user protects himself with no one else's help, and then we review P2P solution, in which users help each other to protect their query privacy.

2.1 Standalone solutions

[15] defines $h(k)$-private information retrieval ($h(k)$-PIR) as a practical compromise between computational efficiency and privacy. They also present $h(k)$-PIR protocols that can be used to query any database, which does not even need to know that the user is trying to preserve his or her privacy. The proposed methods protect the privacy of user queries by adding fake keywords to the keywords really being searched by the user; to prevent the WSE from distinguishing the real from the fake keywords, the latter must be chosen as having a similar frequency of appearance as the former. As a result, the WSE is unable to unequivocally determine the real interests of their users. The quality of the results decreases with the increase in privacy (*i.e* with the number of fake keywords being added), but the trade-off being obtained is excellent.

A prototype called GooPIR was developed in Java JDK 6.0 Standard Edition to implement this scheme (`http://crises2-deim.urv.cat/technology/get/id/1`). The prototype accepts queries consisting of single keywords and queries

consisting of a logical AND of several keywords (with the limitation that independence between the keywords must be a plausible assumption). GooPIR locally masks the target keyword(s), submits the masked query to the Google search engine and then locally filters the results relevant to the target keyword(s).

A standalone solution that was developed in parallel with $h(k)$-PIR by researchers not in ARES is TrackMeNot [23]. Here, instead of adding fake keywords to the real keyword sought by the user, the user's real queries are left unaltered but the system keeps generating and submitting additional fake queries that the WSE cannot easily distinguish from the real ones.

2.2 User-private information retrieval (UPIR)

Like [15, 23], [10, 11] propose to relax strict PIR in order to obtain a practical system. However, rather than altering the user's query with fake queries or cloaking the user's query in a set of queries in a standalone fashion, the user's query history is blurred with the help of a peer-to-peer user community: a user gets her queries submitted on her behalf by other users in the P2P community. In this way, the database still learns which item is being retrieved (which deviates from strict PIR), but it cannot obtain the real query histories of users, which become diffused among the peer users. We name the resulting PIR relaxation user-private information retrieval (UPIR). This approach certainly requires the availability of peers, not needed in standalone systems [15, 23], but it has some advantages: unlike [15], it does not require knowledge of the frequencies of all possible keywords and phrases that can be queried; unlike [23], it avoids the overhead of ghost query submission.

Note that what we offer is different from what can be achieved using anonymization systems based on onion routing, like Tor [44]. In an onion routing system, the transport of data is protected by bouncing the communication between a user and a server around a distributed network of volunteer relays, with a view to protecting against traffic analysis. However, such systems give no end-to-end protection (at the application level). Specifically, as long as a search engine (or a database server) can link the successive queries submitted by the same user (*e.g.* by using cookies or some other mechanism), the profiling and the re-identification capabilities of the search engine are unaffected even if the user is submitting her queries through Tor[1]: the user still submits all of her queries herself (the relays merely relay them), so her query history is unaltered and a query history may suffice for re-identification (see discussion in Section 2.5).

What [10, 11] propose is to diffuse a user's query profile among the peers in a peer-to-peer community. However, onion routing systems can indeed complement our solution and be used for peers to communicate among themselves and hide their identity from each other at the transport level.

The new scheme uses a type of combinatorial design called configuration to manage the keys used by peers to communicate their queries to each other and

[1] However, using the Torbutton browser add-on helps eliminating cookings and allows using Tor to protect the query history of a user.

reduce the number of required keys (see [34, 25] for background on designs and configurations). The use of configurations in cryptographic key management was not new (e.g. see [25]), but their use in private information retrieval was.

2.3 Combinatorial configurations

As indicated in the previous section, configurations are a combinatorial structure playing a central role in UPIR systems. We have done some research on configurations to improve our UPIR protocols.

A first contribution on configurations was [3], followed by [37]. In this latter paper it was proven that the optimal configurations for the P2P UPIR protocol presented in [10, 11] are the finite projective planes. This paper also presented an efficient and explicit algorithm for constructing finite projective planes. Finally, another aspect on the optimality of finite projective planes was treated: a short proof that they are Ramanujan graphs was given.

Subsequent contributions on configurations that were produced in ARES include [4, 38, 39].

2.4 Other collaborative solutions for query profile privacy

In [9] a collaborative approach was presented in which a central node groups users and collects the queries that users want to submit. Then the users execute an anonymous query retrieval protocol and each user obtains from the central node one query without knowing whose query it is. The user submits the query and broadcasts the WSE answer to all other users. This system had the shortcoming that the answer is made public and it might be linkable to the user who originated the query (*e.g.* in case of vanity query this is clear).

In [41] a social network was used for the first time. A new scheme was proposed that was designed to protect the privacy of the users from a web search engine that tries to profile them. The system provides a distorted user profile to the web search engine, because some of each user's queries are submitted by his/her friends in the social network. The proposed protocol submits standard queries to the web search engine, so it does not require any change on the server side. In addition to that, this scheme does not require the server to collaborate with the users.

Nevertheless, the following research questions appear when considering this scheme:

- The privacy level achieved by the users of this proposal depends on the function that calculates the probability of submitting a query. Can this function be re-designed to improve the current results?
- Mechanisms to measure the privacy level achieved by the users are needed in order to compare different proposals. Is there a standard measure that can be used for this purpose?

– The simulations which are shown in [41] have been performed using synthetic queries (queries which are generated at random by a computer). Would the use of real queries (queries generated by humans) influence the behavior of this scheme in terms of privacy protection?

[19] addressed the above research questions:

– The function used to decide which user must submit a certain query to the WSE was studied and re-designed. As a result, the privacy level achieved by the users was improved.
– A new measure to estimate the privacy achieved by the users, the *Profile Exposure Level (PEL)*, was proposed.
– The tests were performed using real data extracted from the well-known AOL file [42]. In this way, the correct behavior of the proposed system was tested with queries which have been generated by real users.

These changes improved the privacy achieved by the users in the previous version, while preserving usability.

Previous proposals of privacy-preserving web search protocols significantly increased the query delay. This is the time that the users need to wait in order to obtain the results of their queries. For this reason, the protocol presented in [29] focused on reducing the query delay. The resulting scheme was implemented and tested in an open environment and the results showed that it achieves the lowest query delay which had been reported in the literature. On the other hand, the work presented in [28] focuses on improving the level of security of previous proposals. More specifically, this work proposed a multi-party protocol that protected the privacy of the user not only in front of the web search engine, but also in front of other members of her own group. The results showed that this scheme outperforms similar proposals in terms of computation and communication.

[8] was developed in collaboration with the Distributed Computation Group of the University of Lleida. This work focused on the development of a P2P network that groups users according to their search preferences. Once the users are classified, they execute a protocol that protects their privacy in front of the web search engine.

2.5 Query log anonymization

The search logs generated by a web search engine are a great source of information for researchers or marketing companies, but at the same time their publication may expose the privacy of the users from which the logs were generated [24]. There is at least one well-known case of released search logs with poor anonymization, which turned out to reveal enough information to re-identify some users. The release was done by AOL in an attempt to help the information retrieval research community, and ended up not only in important damage to the privacy of AOL users, but also in a major damage to AOL itself with several class actions suits and complaints against the company [42]. Ideally, the search logs should be properly anonymized before they become public. The problem is

that achieving an acceptable degree of privacy in search logs is not easy, as there is a trade-off between privacy and the usefulness of the data.

In [17] we presented a method for anonymizing query logs, so as to be able to make them publicly available without encroaching on the privacy of the users who issued the logged queries. To that end, we followed the same ideas found in statistical disclosure control, and proposed a novel microaggregation method to anonymize query logs. This approach ensures a high degree of privacy, and offers k-anonymity at the user level, while preserving some of the data usefulness. Moreover, and unlike most of the previous work, our approach took into account the semantics of the queries in the anonymization process; this was achieved by using the Open Directory Project [43] ontology when aggregating the queries. A more extended version was presented in [18].

Another approach to microaggregating query logs was presented in [26]. In this paper, we defined a new user distance and an aggregation operator. The user aggregation was designed in order to be as computationally efficient as possible. Note that the most important part is the aggregation of the queries, since it is the information that will be most valuable in future analyses. Note also that queries are aggregated separately. An alternative could be to actually mix the terms of queries from different users and end up with new queries that somehow summarize all the users' queries. We chose the former approach given the complexity of the latter, and also because the former method yielded already satisfactory results.

As usual in statistical disclosure control techniques, there is a trade-off between privacy and usability. We showed that our proposals, besides providing k-anonymity, preserve to a good extent the information of the original logs. Our proposals can be regarded as an efficient and relatively simple method to protect query logs, and they ensure a high degree of anonymity and privacy.

3 Location privacy

We will distinguish here between location privacy in location-based services and anonymization of trajectory data for their release.

3.1 Privacy in Location-Based Services

The massive use of mobile devices equipped with self-location technologies such as GPS has fostered the appearance of an unprecedented number of location-based services (LBS) that are gaining importance rapidly. The location-based applications that these new technologies can bring to people are almost unlimited and their advantages very substantial. However, the wide deployment of LBS can jeopardize the privacy of their users and raise social concern. Consequently, ensuring user privacy is essential to the success of those services.

We have mainly focused in TTP-free schemes and collaboration-based methods. [35, 36] refer to approximate location schemes. In [35] the authors proposed a method based on Gaussian noise addition to compute a fake location that is

shared by k users. Thus, all k users use the same fake location and the LBS provider is unable to distinguish one user from the rest, so that their location becomes k-anonymous. This method was extended to support decentralized communications in [36].

On the other hand, [27] presented a new exact location method that has the advantages of these kind of methods such as pseudonymizers (*i.e.* simplicity and accuracy), and avoids their disadvantages (*i.e.* poor scalability and lack of privacy). The idea was to replace the classic concept of pseudonymizer, understood as a TTP, by a distributed pseudonymizer consisting of a set of collaborative users.

3.2 Trajectory anonymization

Trajectories of mobile objects (individuals, cars, etc.) are routinely collected or at least collectible by such technologies as GPS, RFID, GSM, etc. The availability of trajectories, that is, mobility data, is extremely useful for public and corporate planning purposes. However, publication of original collected mobility data would result in obvious privacy disclosure: even if de-identified, trajectories are easily linkable to the individuals they correspond to and they tell a lot about that individual's lifestyle and habits. Furthermore, sensitive locations (hospitals, etc.) visited by individuals may be disclosed.

[14] presented an anonymization method aimed at forming anonymized trajectories with true original locations and providing high utility properties but without a proven privacy level. In [16] the idea of trajectory anonymization by means of location permutation was leveraged, and two new methods were proposed that effectively satisfy provable privacy properties. Moreover, in [14] empirical results were obtained only on synthetic data, while in [16] experiments were added that used a real-life data set of trajectories. Finally, in [40] the formalization of the notion of trajectory k-anonymity given in [16] was used to analyze the privacy offered by (k, δ)-anonymity; it was proven that (k, δ)-anonymity does not offer trajectory k-anonymity when $\delta > 0$, that is, when there is actual uncertainty. A direct implication of this result was that the methods that aimed at achieving (k, δ)-anonymity, *Never Walk Alone* (NWA, [1]) and *Wait for Me* (W4M, [2]) can offer trajectory k-anonymity only when $\delta = 0$ (when there is no uncertainty).

4 Differential privacy

Differential privacy [12, 13] is a statistical disclosure control methodology based on output perturbation.The disclosure risk limitation offered by differential privacy is based on the limitation of the effect that any single individual has on a query response. If the influence of any single individual on the query response is small, publishing that response involves only a small disclosure risk for any individual. The problem with differential privacy is that achieving it normally

results in very damaged data utility. Therefore, the research on differential privacy in ARES set out to find way to satisfy differential privacy that are more utility-preserving than those in the literature.

Any mechanism used to achieve differential privacy may be seen as the application of a perturbation to the real value of the query response. [30] introduced a mechanism to achieve differential privacy that worked by refining the prior knowledge/beliefs of the database user as much as possible, given the constraints set by differential privacy. This mechanism does not require complex computations and it guarantees that the response provides increased utility over the prior knowledge that the user had.

The original proposal [12, 13] to attain differential privacy masked the query response by adding a Laplace distributed noise whose magnitude is proportional to the global sensitivity of the query function. We showed in [31] that the Laplace distribution is not optimal, that is, that differential privacy can be reached with a noise distribution having a lower variance. In that paper, we built the optimal data-independent noise distribution with the help of an optimality criterion based on the concentration of the probability mass of the noise distribution around zero and we compared the resulting distribution with Laplace. For univariate query functions, both introduce a similar level of distortion; however, for multivariate query functions, optimal data-independent noise offers responses with substantially better data quality.

Other ARES contributions to the differential privacy literature highlight synergies between k-anonymity and differential privacy. [32] shows that the amount of noise required to fulfill differential privacy can be reduced if noise is added to a k-anonymous version of the data set, where k-anonymity is reached through a specially designed microaggregation of all attributes. On the other hand, [33] points out that, for data set anonymization, the t-closeness extension of k-anonymity is closely related to differential privacy.

5 Anti-discrimination in data mining

Along with privacy, discrimination avoidance is a very important issue when considering the legal and ethical aspects of data mining. It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, age and so on, especially when those attributes are used for making automated decisions about them like giving them a job, loan, insurance, etc. Discovering such potential biases and eliminating them from the data used to train data mining classifiers without harming their decision-making utility is therefore highly desirable. For this reason, anti-discrimination techniques including discrimination discovery and prevention have been introduced in data mining.

Discrimination can be either direct or indirect. Direct discrimination occurs when decisions are made based on sensitive attributes. Indirect discrimination occurs when decisions are made based on nonsensitive attributes which are strongly correlated with biased sensitive ones.

In a first work [20], we introduced the initial idea of using rule protection and rule generalization for direct discrimination prevention, but we gave no experimental results. In [21], we introduced the use of rule protection in a different way for indirect discrimination prevention and we gave some preliminary experimental results. Finally, [22] presented a unified approach to direct and indirect discrimination prevention, with finalized algorithms and all possible data transformation methods based on rule protection and/or rule generalization that could be applied for direct or indirect discrimination prevention.

6 Conclusions

The overview that we have presented in this chapter is intended as a reading guide to the some of the contributions of ARES to data privacy technologies related to query profile protection, privacy in location-based systems, trajectory anonymization, differential privacy and anti-discrimination. Further details can be obtained by looking at the corresponding publications or at the other chapters in this book.

Acknowledgments and disclaimer

The second author is partially supported by the Government of Catalonia through an ICREA Acadèmia Prize. The following partial supports are also gratefully acknowledged: the Spanish Government under projects CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES" and TIN2011-27076-C03-01 "CO-PRIVACY", and the European Commission under FP7 projects "DwB" and "Inter-Trust". The second author is with the UNESCO Chair in Data Privacy, but the views expressed in this chapter neither necessarily reflect the position of UNESCO nor commit that organization.

References

1. O. Abul, F. Bonchi, M. Nanni: Never walk alone: uncertainty for anonymity in moving objects databases, in: 24th International Conference on Data Engineering, pp. 376-385 (2008)
2. O. Abul, F. Bonchi, M. Nanni: Anonymization of moving objects databases by clustering and perturbation. Inf. Syst. 35 (8), pp. 884-910 (2010).
3. M. Bras-Amorós, J. Domingo-Ferrer, K. Stokes: Configuraciones combinatóricas y recuperación privada de información por pares. In: Nuevos Avances en Criptografía y Codificación de la Información (2009).
4. M. Bras-Amorós, K.Stokes: The semigroup of combinatorial configurations. Semigroup Forum 84 (1), pp. 91-96 (2011)
5. B. Chor, O. Goldreich, E. Kushilevitz, M. Sudan: Private information retrieval. In: IEEE Symposium on Foundations of Computer Science, pp. 41-50 (1995)
6. B. Chor, N. Gilboa, M. Naor: Private Information Retrieval by keywords. Technical Report TR CS0917, Departament of computer Science, Technion (1997).

7. B. Chor, O. Goldreich, E. Kushilevitz, M. Sudan: Private information retrieval. Journal of the ACM 45, pp. 965-981 (1998)
8. D. Castellà, C. Romero-Tris, A. Viejo, J. Castellà-Roca, F. Solsona, F. Giné: Diseño de una red P2P optimizada para la privatización de consultas en WSEs. In: XII Reunión Española sobre Criptología y Seguridad de la Información, pp. 273-278 (2012)
9. J. Castellà-Roca, A. Viejo, J. Herrera-Joancomartí: Preserving users' privacy in web search engines. Computer Communications 32 (13), pp. 1541-1551 (2009)
10. J. Domingo-Ferrer, M. Bras-Amorós: Peer-to-peer private information retrieval. In: PSD 2008. LNCS, vol. 5262, pp. 315-323 (2008)
11. J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, J. Manjón: User-Private Information Retrieval Based on a Peer-to-Peer Community. Data & Knowledge Engineering 68 (11), pp. 1237-1252 (2009)
12. C. Dwork, F. McSherry, K. Nissim, and A. Smith: Calibrating noise to sensitivity in private data analysis. In Proceedings of the 3rd Theory of Cryptography Conference. LNCS, vol. 3876, pp 265-284. Springer (2006).
13. C. Dwork: Differential privacy. In Automata, Languages and Programming. LNCS, vol. 4052, pp 1-12. Springer (2006)
14. J. Domingo-Ferrer, M. Sramka, R. Trujillo: Privacy preserving Publication of Trajectories Using Microaggregation. In 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS (2010).
15. J. Domingo-Ferrer, A. Solanas, J. Castellà-Roca: $h(k)$-private information retrieval from privacy-uncooperative queryable databases. Online Information Review 33 (4), pp. 720-744 (2009)
16. J. Domingo-Ferrer, R. Trujillo-Rasua: Microaggregation- and permutation-based anonymization of movement data. Information Sciences 208, pp. 55-80 (2012)
17. A. Erola, J. Castellà-Roca, G. Navarro-Arribas, V. Torra: Semantic Microaggregation for the Anonymization of Query Logs. In: PSD 2010. LNCS, vol. 6344, pp. 127-137 (2010)
18. A. Erola, J. Castellà-Roca, G. Navarro-Arribas, V. Torra: Semantic microaggregation for the anonymization of query logs using the open directory project. SORT-Statistics and Operations Research Transactions, Special issue, pp. 41-58 (2011)
19. A. Erola, J. Castellà-Roca, A. Viejo, J.M. Mateo-Sanz: Exploting Social Networks to Provide Privacy in Personalized Web Search. Journal of Systems and Software 84 (10), pp. 1734-1745 (2011)
20. S. Hajian, J. Domingo-Ferrer, A. Martínez-Ballesté: Discrimination prevention in data mining for intrusion and crime detection. In: IEEE Symposium Series in Computational Intelligence in Cyber Security (2011).
21. S. Hajian, J. Domingo-Ferrer, A. Martínez-Ballesté: Rule protection for indirect discrimination prevention in data mining. In: MDAI 2011. LNCS, vol. 6820, pp. 211-222 (2011)
22. S. Hajian, J. Domingo-Ferrer: A methodology for direct and indirect discrimination prevention in data mining. IEEE Transactions on Knowledge and Data Engineering 25 (7), pp. 1445-1459 (2013)
23. D.C. Howe, H. Nissenbaum: TrackMeNot: resisting surveillance in web search. In: I. Kerr, C. Lucock, V. Steeves (Eds.), Lessons from the Identity Trail: Privacy, Anonymity and Identity in a Networked Society, Oxford University Press, Oxford UK, pp. 409-428 (2009)
24. R. Jones, R. Kumar, B. Pang, A. Tomkins: I know what you did last summer: Query logs and user privacy. In Proceedings of the sixteenth ACM conference on conference on information and knowledge management, pp. 909-914 (2007)

25. J. Lee, D.R. Stinson: A combinatorial approach to key predistribution for distributed sensor networks. In: Wireless Communications and Networking Conference-WCNC 2005, vol. 2, pp. 1200-1205 (2005)
26. G. Navarro-Arribas, V. Torra, A. Erola, J. Castellà-Roca: User k-anonymity for privacy preserving data mining of query logs. Information Processing and Management 48 (3), pp. 476-487 (2012)
27. P. A. Pérez-Martínez, A. Solanas: Location Privacy Through Users' Collaboration: A Distributed Pseudonymizer. In: Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (2009).
28. C. Romero-Tris, J. Castellà-Roca, A. Viejo: Multi-party private web search with untrusted partners. In: 7th International Conference on Security and Privacy in Communication Networks (2011)
29. C. Romero-Tris, A. Viejo, J. Castellà-Roca: Improving query delay in private web search. In: International Workshop on Securing Information in Distributed Environments and Ubiquitous Systems (2011)
30. J. Soria-Comas, J. Domingo-Ferrer: Sensitivity-Independent Differential Privacy via Prior Knowledge Refinement. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 20 (6), pp. 855-876 (2012)
31. J. Soria-Comas, J. Domingo-Ferrer: Optimal data-independent noise for differential privacy. Information Sciences 250, pp. 200-214 (2013)
32. J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, S. Martínez: Improving the utility of differentially private data releases via k-anonymity. In: 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (2013).
33. J. Soria-Comas, J. Domingo-Ferrer: Differential privacy via t-closeness in data publishing. In: 11th Annual Conference on Privacy, Security and Trust, pp. 27-35 (2013).
34. D.R. Stinson: Combinatorial Designs: Constructions and Analysis. SIGACT News 39(4), pp. 17-21 (2008)
35. A. Solanas, A. Martínez-Ballesté: Privacy protection in location-based services through a public-key privacy homomorphism. In: Euro PKI 2007. LNCS, vol. 4582, pp. 362-368 (2007)
36. A. Solanas, A. Martínez-Ballesté: A TTP-free protocol for location privacy in location-based services. Computer Communications 31 (6) pp. 1181-1191 (2008).
37. K. Stokes, M. Bras-Amorós: Optimal Configurations for Peer-to-Peer User-Private Information Retrieval. Computers & Mathematics with Applications 59 (4), pp. 1568-1577 (2010)
38. K. Stokes, M. Bras-Amorós: Associating a numerical semigroup to the triangle-free configurations. Advances in Mathemathics of Communication 5 (2), pp. 351-371 (2011)
39. K.Stokes, O.Farràs: Linear spaces and transversal designs: k-anonymous combinatorial configurations for anonymous database search. Designs, Codes and Cryptography 71, pp. 503-524 (2014)
40. R. Trujillo, J. Domingo-Ferrer: On the privacy offered by k-d-anonymity. Information Systems 38 (4), pp. 491-494 (2013)
41. A. Viejo, J. Castellà-Roca: Using Social Networks to Distort Users' Profiles Generated by Web Search Engines. Computer Networks 54 (9), pp. 1343-1357 (2010)
42. AOL Search Data Scandal, August 2006. `http://en.wikipedia.org/wiki/AOL_search_data_leak`
43. ODP. Open directory project. `http://www.dmoz.org/`

44. The Tor Project, Inc: Tor: Overview. http://torproject.org/overview.html.en