The Functionality-Security-Privacy Game

Josep Domingo-Ferrer*

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

Abstract. Privacy preservation in the information society is in many respects parallel to environment preservation in the physical world. In this way, "green ICT services" are those achieving functionality and security with minimum invasion of the privacy of individuals, where such an invasion can be regarded as a kind of pollution as harmful in the long run to their moral welfare as physical pollution is to their physical welfare. Depending on the type of service, individuals can be users, data owners or respondents having supplied data. We show that the conflict between functionality, security and privacy can be viewed as a game between several players whose interests differ. If the game is properly formulated, its equilibria can lead to protocols conciliating the functionality, security and privacy interests of all players.

Keywords: Privacy, Security, Functionality, Game theory, Mechanism design.

1 Introduction

The starting point of this paper is that *privacy preservation in the information society is analogous to environment preservation* in the physical world. With this idea in mind, "green" or "clean" information and communications technologies (ICT) are those offering functionality and security with minimum invasion of the privacy of individuals. Such an invasion can be regarded as a virtual pollution as harmful in the long run to the moral welfare of individuals as physical pollution is to their physical welfare. The moral value of privacy was of course previous to the information society: privacy is a fundamental right of the individual, acclaimed by the United Nations in article 12 of the Universal Declaration of Human Rights (1948). In fact, the lack of privacy undermines most of the remaining fundamental rights (freedom of speech, democracy, etc.).

* The author is with the UNESCO Chair in Data Privacy, but the views expressed in this paper are those of the author and do not commit UNESCO. This work was partly funded by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS" and CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES", and by the Government of Catalonia through grant 2009 SGR 1135. The author is partly supported as an ICREA-Acadèmia researcher by the Government of Catalonia.

V. Torra, Y. Narukawa, and M. Inuiguchi (Eds.): MDAI 2009, LNAI 5861, pp. 92–101, 2009.

1.1 The Pollutants of Privacy

One source of privacy pollution has to do with *privacy-unfriendly security*. Security in general and information security in particular are fields of increasing relevance. The advent of R+D funding priorities related to security is an indication and a confirmation of the previous statement: the European Union's 7th Framework Programme (FP7) includes, among its ten thematic priorities, one named "Security". Additionally, security and privacy are nearly ubiquitous in a second FP7 thematic priority among those ten, the one named "Information and Communication Technologies (ICT)". Especially focused to security is the first of the seven challenges addressed by this priority, named "Pervasive and Trusted Network and Service Infrastructures".

This interest of governments and corporations in security is partly justified by the social alarm in front of the global threat of international terrorism. With the argument of such a threat, the European Union and several overseas states have adopted shock measures on information security. Beyond the obvious technological challenge of securing and analyzing communications on a mass scale, a new, subtler and unaddressed challenge arises: security must be increased while minimizing the loss of privacy for the citizens. This second challenge becomes especially pressing after the measures adopted in Europe about keeping track of phone calls and e-mail messages. The tendency of governments is to sacrifice privacy for security (*e.g.* the former UK security and intelligence coordinator recently asserted that anti-terror fight will need privacy sacrifice [15]). Similar conflicts between privacy and security appear in connection with the identity theft in bank transactions by the organized crime [24]. The attraction of focusing on security technologies while putting aside privacy and other rights of the individuals is very strong as it can be inferred even from the FP7 stance about privacy, which is mentioned as an ethical-legal issue rather than as technological objective *per se*. In general, increasing security without significantly decreasing privacy is one of the main challenges faced by the information society.

A second source of privacy pollution is *privacy-unaware* (let alone privacy-unfriendly) *functionality*. Many ICT services, like search engines (*e.g.* Google, Yahoo, etc.), social networks and most Web 2.0 services, the constellation of Google additional services (Calendar, Streetview, Latitude), etc., concentrate on offering enticing functionality for users while completely disregarding their privacy. They are like powerful cars which pollute a lot. Whenever one of such services boasts a privacy pledge, it refers to privacy in front of third parties (*e.g.* the service provider commits to abstaining from unauthorized transfer of user data to third parties) but not to privacy of the user in front of the service provider itself (*e.g.* Google learns and records all queries submitted to its search engine, all events entered in the Calendar, all e-mails sent/received with Gmail, etc.). Hence, each service provider (Google, Yahoo, Facebook, etc.) becomes a big brother in the purest Orwellian sense.

1.2 Contribution and Plan of This Paper

We show that the conflict between functionality, security and privacy can be expressed as a game between some players, whose number, nature and utility functions depend on the specific application scenario. If the game is designed to deter rational deviations by the players, its equilibria can lead to protocols conciliating the functionality, security and privacy interests of the players.

Section 2 states the general game-theoretic approach. Section 3 illustrates the approach in several application scenarios, namely, statistical disclosure control, car-to-car communication, private information retrieval and social networks. Conclusions and lines for future research are summarized in Section 4.

2 A Game-Theoretic Framework

The trade-off between privacy, security and functionality can be viewed as a game [23, 19, 21] between an individual and a system (which may be an organization, a computer system or an ICT service):

- The individual wishes to obtain *functionality* with minimum *privacy* loss. Think of an e-commerce transaction by way of illustration: regarding functionality, the buyer expects the electronic shop to have a good catalog and be convenient to use; regarding privacy, the buyer wants to pay without her credit card number being stolen, and she wants the system to keep her purchase record confidential or, even better, not to keep any record about her at all, unless she is offered some rewards, like improved customer relationship or discounts.
- The system's primary goals are *functionality* and *security* of its own strategic information (accounting, inventory levels, digital content if the system trades with information as a commodity, etc.); this kind of security could also be termed system privacy as opposed to privacy of individuals, but we stick to the term security to avoid confusion. Individual privacy (related to customers or subjects the system holds information about) is, at best, a secondary goal. In the e-commerce example, the system is the e-shop, which wants to offer functionality to customers, while keeping its backoffice information secure and confidential (stocks, sales, etc.); regarding customers, the usual aim of the e-shop is to profile them as much as possible in view of improving the customer relationship management.

Hence, functionality is a goal shared by the individual and the system, but privacy and security are not. We have so far described the functionality-security-privacy game as one between two players: the individual user and the system. However, when the system holds data about third parties, *e.g.* if the system is a database holding records about individual respondents, then those respondents are a third player (or a community of "third players"), whose main goal is to preserve their privacy; the information released by the system to individual users should not be linkable to specific individual respondents. Key questions whose answer could be obtained with such a game-theoretic framework include the following:

1. How much privacy is the individual user willing to surrender in exchange for increased functionality? How much functionality can be offered for a certain level of user privacy?
2. Can the system offer the required functionality and user privacy while guaranteeing its own security (and the privacy of respondents, if there are respondents)?

In [8], this author presented a three-dimensional framework for privacy in databases. The three dimensions correspond to the three types of privacy sought: user privacy, system privacy (called above system security) and respondent privacy. That paper showed the independent nature of those types of privacy and the need to trade them off; it also assessed how well various technologies for database privacy managed to provide the three types of privacy; the comparison included privacy information retrieval (PIR, [5]), noise-based privacy-preserving data mining (noise-based PPDM, [1]), cryptographic privacy-preserving data mining (crypto PPDM, [17]) and statistical disclosure control (SDC, [14]). The main limitations of [8] were the following: i) only database privacy was discussed, but not privacy in other ICT applications; ii) functionality was not considered; and iii) the assessment of technologies was qualitative and lacked quantitative criteria.

Progressing to a general game-theoretic framework requires going through the steps below:

1. If there are n players, $\{1, 2, \cdots, n\}$ in the functionality-security-privacy game, identify the set of possible strategies S_i of each player i. If player i selects a strategy $s_i \in S_i$, denote by $s = (s_1, \cdots, s_n)$ the vector of strategies selected by the players.
2. For each player i, find functionality metrics $f_i(s)$, security metrics $sec_i(s)$ and privacy metrics $p_i(s)$. Such metrics will be application-dependent, as argued below.
3. For each player i, find utility functions $u_i(s)$ mixing the above functionality, security and privacy metrics. The utility for the individual user is likely to be a mixture of functionality and privacy; for the system, it is likely to be a mixture of functionality and security; for the respondent, it is likely to be basically privacy. Finding the optimal mixture function is in itself a decision-theoretic problem faced by each player. If we denote by s_i the strategy played by player i and by s_{-i} the $(n-1)$-dimensional vector of the strategies played by the other players, we can write $s = (s_i, s_{-i})$. Whatever the utility mixture, if s_i' is an alternate strategy to s_i, it should hold that:

$$u_i(s_i, s_{-i}) = u_i(s_i', s_{-i}) \text{ if } f_i(s_i) = f_i(s_i') \text{ and } sec_i(s_i) = sec_i(s_i') \text{ and } p(s_i) = p(s_i')$$

$$u_i(s_i, s_{-i}) > u_i(s_i', s_{-i}) \text{ if } \begin{cases} f_i(s_i) > f_i(s_i') \text{ and } sec_i(s_i) \geq sec_i(s_i') \text{ and } p(s_i) \geq p(s_i') \\ f_i(s_i) \geq f_i(s_i') \text{ and } sec_i(s_i) > sec_i(s_i') \text{ and } p(s_i) \geq p(s_i') \\ f_i(s_i) \geq f_i(s_i') \text{ and } sec_i(s_i) \geq sec_i(s_i') \text{ and } p(s_i) > p(s_i') \end{cases}$$

4. Use mechanism design [19] to find game rules which, combined with the players' utility functions, ensure that:
   (a) Players will not deviate from the prescribed rules;
   (b) The game results in equilibria acceptable to all players.

3 Specific Application Scenarios

For the sake of concreteness, we discuss the above game-theoretic framework in several specific application scenarios, namely: statistical disclosure control, user-private information retrieval, car-to-car communication and social networks.

3.1 Statistical Disclosure Control

Statistical disclosure control (shortened as SDC and known as Privacy-Preserving Data Mining in the database community, see *e.g.* [10]) seeks to protect statistical data in such a way that they can be publicly released and mined by users without giving away private information which can be linked to specific individuals or entities (the respondents who supplied the data). This is a case of functionality-security-privacy game, which was partially stated for the specific case of tabular data in [18].

One challenge is to model SDC as a functionality-security-privacy game for any kind of data, which requires quantifying risks and pay-offs in order to construct suitable utility functions. Players are the database owner, users and respondents:

- The owner (typically a national statistical office) wants security, that is, to make sure that no data misuse will occur; hence, the owner's utility function u_o is proportional to the probability of users *not* misbehaving.
- The respondent wants privacy, that is, guarantees that no user or intruder will be able to link her responses with her identity; hence, the respondent's utility function u_r is proportional to the disclosure risk in the published data set.
- The user wants functionality, that is, maximum analytic flexibility and accuracy; hence, the user's utility function u_u is inversely proportional to the information loss caused by the anonymization process (from the user's point of view, the best anonymization is just releasing the original data unaltered).

Thus, SDC turns out to be a case in which the utilities of the players are "pure", that is, each player type is interested in one and only one property. However, at a closer look, the owner's and the respondent's utility do have some correlation, because they are both maximized when the user is "under control": if disclosure risk is low, the user has less chances to misbehave.

The strategies of the SDC game can be summarized as: i) for the data owner, the anonymization procedure and the computer security defenses chosen; ii) for the respondent, whether to respond accurately, to respond inaccurately or not to respond at all; iii) for the user, whether to make use of the anonymized data (and maybe pay for them) or to reject the anonymized data.

The utility functions should be such that it is in each player's own interest to behave rationally in order to maximize their utility. In this case, the equilibria of the game would yield parameterizations of the anonymization procedures which optimize the trade-off between functionality, security and privacy, that is, between information loss, security and disclosure risk.

3.2 User-Private Information Retrieval

Private information retrieval (PIR) is normally modeled as a game between two players: a user and a database. The user retrieves some item from the database without the latter learning which item was retrieved. Most current PIR protocols are ill-suited to provide PIR from a search engine or large database: i) their computational complexity is linear in the size of the database; ii) they assume active cooperation by the database in the PIR protocol. If the database cannot be assumed to cooperate, two pragmatic approaches can be adopted by the user to keep her query interests private:

1. *Standalone.* A program in the user's computer keeps submitting fake queries to the search engine at random times, so that when the user submits a real query, the search engine cannot distinguish it from the fake queries; this is the TrackMeNot approach [13]. An alternative standalone approach which is less resource-consuming is for the user to mask her real query keywords by adding some fake keywords with similar frequency; in this way, the number of queries submitted equals the number of real user queries; this is the way our GooPIR prototype operates [11].
2. *Peer-to-peer (P2P).* 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, but it cannot obtain the real query histories of users, which become diffused among the peer users. We named this relaxation user-private information retrieval and published it in [12].

PIR and its standalone and P2P relaxations can be modeled as functionality-privacy games. The user wants to query the search engine or database with as much flexibility and speed as possible (functionality), while keeping her query history private (privacy). In fact, the main problem of strict PIR is that, while it achieves very good privacy, it offers only very restricted functionality. On the other hand, the P2P relaxation is a game with several players: the peers and the database, and user privacy means privacy in front of the database and in front of the rest of peer users. Determining the optimal values of parameters such as the number of peers and the connectivity among peers can be an outcome of this modeling process.

A further challenge is to design a new relaxation of PIR offering also system security. If the database information items are not free and have different prices, system security means that users should not be able to retrieve items without paying the established fees. Thus, the game should not be just functionality-privacy, but functionality-security-privacy.

3.3 Car-to-Car Communication

Vehicular *ad hoc* networks (VANETs) allowing car-to-car communication are expected to be commercially available in cars manufactured from 2011 onwards [3]. There are several standards for VANET communication under way both in the United States (DSRC, Dedicated Short Range Communications, IEEE 802.11p)

and Europe (Car2Car Consortium). VANETs will allow vehicles to disseminate announcement messages about road conditions to other vehicles (icy road, traffic jam, etc.), in order to improve traffic safety and efficiency.

Security is a clear requirement of such announcement messages: these must be trustworthy, because false messages could seriously disrupt traffic, cause accidents or cause certain areas to become deserted and hence easy targets for criminals. There are two approaches in the literature: *a posteriori* and *a priori*. Under the *a posteriori* approach, punitive action is taken against vehicles which have been proven to have originated false messages (*e.g.* [16]); hence, means are required to identify malicious vehicles. Under the *a priori* approach, the goal is to prevent the generation of false messages (*e.g.* [20]): a message is given credit only if it is endorsed by a number of nearby vehicles greater than a certain threshold.

Privacy is also a requirement in VANET communication, although perhaps less compelling for carmakers and policy makers. Indeed, it would not be very fair if the collaboration of a driver to improve traffic safety and efficiency by generating or endorsing announcements forced her to disclose her identity and location. Note that knowing the mobility pattern of someone reveals a lot of private information: the way of driving tells a lot about an individual's character (nervous, calm, etc.), her whereabouts give information on her work and social habits, etc. *Thus VANET communication is a case of the functionality-security-privacy game* mentioned above: functionality for individual cars and the overall traffic system, security for the traffic system and privacy for the individual cars.

Privacy can be added to the *a posteriori* security approach by using pseudonyms or more advanced cryptographic primitives like group signatures. A trusted third party is needed who can open the identities of honest vehicles.

Adding vehicle privacy to the *a priori* security approach can imply vulnerability against the Sybil attack, in which a vehicle generates a false message and takes advantage of anonymity to compute itself as many endorsements as required. A private *a priori* scheme for VANET announcements based on threshold signatures and resistant against the Sybil attack was recently proposed in [6]. In that paper, irrevocable anonymity for cars generating or endorsing messages is provided.

A posteriori countermeasures alone are not sufficient. They can indeed deter some rational attackers, but they cannot prevent damage by irrational attackers (*e.g.* terrorists willing to risk anything to disseminate false messages aimed at causing an accident or a massive jam). On the other hand, *a priori* countermeasures alone are not sufficient either: if a number of nearby attackers greater than the preset threshold collude, they can generate a valid false message. If they enjoy irrevocable anonymity, those colluders cannot even be traced.

The above shortcomings have been recently addressed in [25] by:

- Designing new group signatures such that: i) they offer anonymity revocability; ii) signatures on different messages by different signers are indistinguishable; iii) signatures on the same message by different signers are distinguishable (so that the Sybil attack can be detected); iv) they are computationally efficient.

– Based on the above signatures, giving a solution to the VANET functionality-privacy-security game offering both *a priori* and *a posteriori* security and privacy for honest vehicles. A traffic load-dependent threshold for *a priori* security and a trusted third party to handle revocation in the case of a posteriori security are used. The solution allows finding the optimal threshold for given traffic conditions.

3.4 Social Networks

Social networks have become an important web service with a broad range of applications: collaborative work, collaborative service rating, resource sharing, friend search, etc. Facebook, MySpace, Xing, etc. are well-known examples. In a social network, a user publishes and shares information and services. In some social networks, the user can specify how much it trusts other users, by assigning them a trust level. It is also possible to establish several types of relationships among users (like "colleague of", "friend of", etc.). The trust level and the type of a relationship are used to decide whether access is granted to resources and services being offered (*access rule*).

The availability of information on relationships (trust level, relationship type) has increased with the advent of the Semantic Web and raises privacy concerns: knowing who is trusted by a user and to what extent discloses a lot about the user's thoughts and feelings. For a list of related abuses see [2]. Also, it is known that some human resources departments use to look up job applicants in a well-known and privacy-weak social network like Facebook to find out more about their personality. Hence, *social networks are another instance of the functionality-security-privacy game*; functionality means resource availability and flexibility of access, which should be possible even if there are only indirect relationships between the resource owner and the resource requestor (as considered in [4]); security refers to the resource owner, who wants to make sure that her resource will be accessed only by users whom she trusts enough; privacy refers to users, who should be able to access resources (or help other users in accessing them) with minimum disclosure of the identities of their relationships.

In [9], a new protocol was described which offers private relationships in a social network while allowing resource access through indirect relationships without requiring a mediating trusted third party. Thanks to homomorphic encryption, this scheme prevents the resource owner from learning the relationships and trust levels between the users who collaborate in the resource access (these users are intermediate relationships between the resource owner and the resource requestor). In this way, the number of users who might refuse collaboration due to privacy concerns is minimized. This results in increased functionality, *i.e.* availability.

Research avenues in social networks are:

– State the problem of resource access via indirect private relationships as a functionality-security-privacy game.
– Use that game-theoretic formulation to design access rules which optimize the pay-offs of users in terms of functionality, security and privacy.

4 Conclusions and Future Research

We have presented the conflict between functionality, security and privacy as a game, and we have illustrated the possible ramifications of such a game-theoretic framework in a number of specific application scenarios.

Future research will involve: i) turning those ramifications into concrete results for each scenario by addressing the challenges identified above; and ii) tackling scenarios not considered here. This requires specifying utility functions and strategies for players.

A final caution is in order, though. As pointed out in [7], there are some fundamental differences between game theory, where players are not supposed to deviate from the game rules, and cryptographic or security protocols, where deviation must be accounted for. Such a shortcoming can be sometimes mitigated with a careful design of the game mechanism (the game rules), *i.e.* so that it is in the players' own interest not to deviate (take for example the Vickrey auction mechanism, [22]). However, for some practical application scenarios, no game mechanism may exist which discourages all possible player deviations: such is the case in the scenario of secure multiparty computation considered in [19], where the class of non-cooperatively computable functions currently seems to be the only one whose computation can be modeled as a game.

References

1. Agrawal, R., Srikant, R.: Privacy preserving data mining. In: Proceedings of the ACM SIGMOD, pp. 439–450 (2000)
2. Barnes, S.B.: A privacy paradox: social networking in the United States. First Monday 11(9) (2006)
3. Blau, J.: Car talk. IEEE Spectrum 45(10), 16
4. Carminati, B., Ferrari, E.: Private relationships in social networks. In: Private Data Management PDM 2007. IEEE Press, Los Alamitos (2007)
5. Chor, B., Goldreich, O., Kushilevitz, E., Sudan, M.: Private information retrieval. In: IEEE Symposium on Foundations of Computer Science (FOCS), pp. 41–50 (1995)
6. Daza, V., Domingo-Ferrer, J., Sebé, F., Viejo, A.: Trustworthy privacy-preserving car-generated announcements in vehicular *ad hoc* networks. IEEE Transactions on Vehicular Technology 58(4), 1876–1886 (2009)
7. Dodis, Y., Rabin, T.: Cryptography and game theory. In: [19], pp. 181–205
8. Domingo-Ferrer, J.: A three-dimensional conceptual framework for database privacy. In: Jonker, W., Petković, M. (eds.) SDM 2007. LNCS, vol. 4721, pp. 193–202. Springer, Heidelberg (2007)
9. Domingo-Ferrer, J., Viejo, A., Sebé, F., González-Nicolás, Ú.: Privacy homomorphisms for social networks with private relationships. Computer Networks 52, 3007–3016 (2008)
10. Domingo-Ferrer, J.: A survey of inference control methods for privacy-preserving data mining. In: Aggarwal, C., Yu, P. (eds.) Privacy-Preserving Data Mining: Models and Algorithms. Advances in Database Systems, vol. 34, pp. 53–80. Springer, Heidelberg (2008)

11. Domingo-Ferrer, J., Solanas, A., Castellà-Roca, J.: $h(k)$-Private information retrieval from privacy-uncooperative queryable databases. Online Information Review 33(4), 720–744 (2009)
12. Domingo-Ferrer, J., Bras-Amorós, M., Wu, Q., Manjón, J.: User-private information retrieval based on a peer-to-peer community. Data and Knowledge Engineering (in press, available online doi:10.1016/j.datak.2009.06.004)
13. Howe, D.C., Nissenbaum, H.: TrackMeNot: Resisting surveillance in web search. In: Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society, pp. 409–428. Oxford University Press, Oxford (2009)
14. Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Longhurst, J., Schulte-Nordholt, E., Seri, G., DeWolf, P.-P.: Handbook on Statistical Disclosure Control (version 1.0), Eurostat (CENEX SDC Project Deliverable) (2006)
15. Hutt, R., Omand, D.: Anti-terror fight will need privacy sacrifice. In: The Independent, February 25 (2009)
16. Lin, X., Sun, X., Ho, P.H., Shen, X.: GSIS: A secure and privacy preserving protocol for vehicular communications. IEEE Transactions on Vehicular Communications 56(6), 3442–3456 (2007)
17. Lindell, Y., Pinkas, B.: Privacy-preserving data mining. Journal of Cryptology 15(3), 177–206 (2002)
18. Mackey, E.: A Framework for Understanding Statistical Disclosure Processes, Ph. D. Thesis, University of Manchester (2009)
19. Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V.: Algorithmic Game Theory. Cambridge University Press, Cambridge (2007)
20. Raya, M., Aziz, A., Hubaux, J.-P.: Efficient secure aggregation in VANETs. In: Proc. of 3rd Intl. Workshop on Vehicular Ad Hoc Networks-VANET, pp. 67–75.
21. Shoham, Y., Leyton-Brown, K.: Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press, Cambridge (2009)
22. Tardos, É., Vazirani, V.V.: Basic solution concepts and computational issues. In: [19], pp. 3–28
23. Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
24. The Wall Street Journal, 34 (December 1, 2005)
25. Wu, Q., Domingo-Ferrer, J., González-Nicolás, Ú.: Balanced trustworthiness, safety and privacy in vehicle-to-vehicle communications (manuscript, 2009)