Co-utilitat: equilibris racionals de privadesa, seguretat i funcionalitat a la societat de la informació

Josep Domingo-Ferrer

Universitat Rovira i Virgili, Tarragona

Càtedra de Privades @ de dades

josep.domingo@urv.cat

1. Introduction
2. Basics of game theory
3. Co-utility
   - The concept
   - Types of equilibria in co-utility
4. Co-utility example applications
   - Anonymous keyword search
   - Social networks
   - Fingerprinted multicast and digital oblivion
5. Cryptographic primitives for gradual privacy/utility
6. Conclusions and research directions
7. Related papers

Introduction: an environmental analogy

- Privacy preservation is essential to make the information society sustainable just as environment preservation is essential to make the physical world sustainable. Privacy invasion is a virtual pollution as harmful to the moral welfare of individuals as physical pollution is to their physical welfare.
- Privacy preservation itself should be sustainable and be achieved as effortlessly as possible as the result of rational co-operation rather than as an expensive legal requirement. Otherwise, privacy will not be global.

Pollutants of privacy

- Privacy-unfriendly security: sacrifice privacy with the excuse of security (*e.g.* anti-terror fight at the expense of privacy, biometrics enforced on customers with the argument of identity theft fighting).
- Privacy-unaware functionality. Enticing functionality offered to users while disregarding their privacy, like in social networks, search engines and Web 2.0 services (*e.g.* Google Calendar, Streetview, Latitude). If there is privacy, it is vs third parties, not vs the provider.

The three "R": reducing, reusing and recycling

Reducing

Re-identifiable information must be reduced. Reducing the informational content of quasi-identifiers is precisely the goal of k-anonymization via recoding or microaggregation. Reduction is also behind ring and group signatures: signer identifiability is reduced.

- Limits to information reduction. *E.g.* eliminating quasi-identifiers dramatically reduces data utility (functionality problem) and deleting the signature in a message suppresses authentication (security problem).
- Privacy is gradual. Privacy preservation is not all-or-nothing, it is a continuous magnitude from no privacy to full privacy preservation.

The three "R": reducing, reusing and recycling

Reusing

Reusing is in the mind of impersonators mounting replay attacks, but it can also be used to gain privacy:

- Limits to reusing. The more reuse, the less data utility.
- Resampling is reusing. An original data set with N records is re-sampled M times with replacement (where M can be even greater than N) and the resulting data set with M records is released instead of the original one. This idea is behind synthetic data generation via multiple imputation.

The three "R": reducing, reusing and recycling

Recycling

It can be regarded as leveraging other people's efforts to preserve their privacy to preserve one's own privacy.

- Limits to recycling. One must adjust to the needs of other people.
- Potential of recycling. Privacy becomes an attractive and shared goal, and is thus easier to achieve and more sustainable. This the idea behind co-utility and specifically co-privacy.

Game theory

- A game is a protocol between N players $\{P^1, \cdots, P^N\}$.
- Each player P^i has her own *set of possible strategies*, say S_i.
- Each player P^i selects a strategy $s_i \in S_i$ and let $s = (s_1, \cdots, s_N)$ denote the vector of strategies selected by players.
- If $S = \Pi_i S_i$ is the set of all possible ways in which players can pick strategies, define the utility of P^i as $u_i : S \longrightarrow \mathbb{R}$.
- For $s \in S$, denote by s_i the strategy chosen by P^i and s_{-i} the strategies played by all other players.

Dominant strategies and Nash equilibria

Dominant strategy

It is s such that for each player P^i and each alternate strategy vector $s' \in S$, $u_i(s_i, s'_{-i}) \geq u_i(s'_i, s'_{-i})$

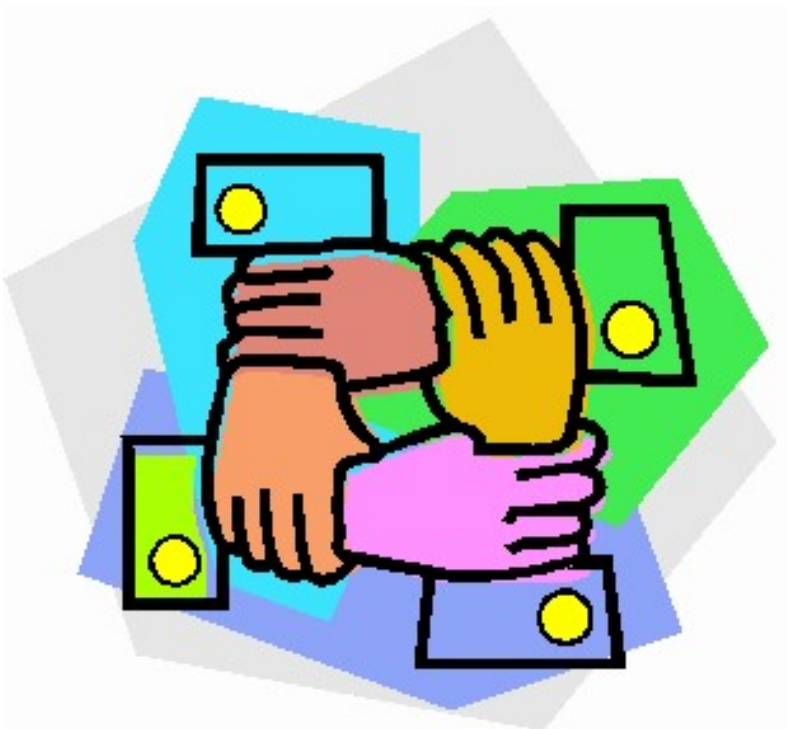
Nash equilibrium

It is s such that for each player P^i and each alternate strategy $s'_i \in S_i$, it holds that $u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$

A dominant strategy is best no matter what others do. A Nash equilibrium is best if all other players stick to their equilibrium strategies.

The concept of co-utility

There is co-utility in a community of players when the best way for a player to serve her own interests is to help one or more players in their interests.



Co-utility vs co-privacy

- Co-utility generalizes the co-privacy concept introduced in Domingo-Ferrer (2010 and 2011).
- If players' interests include privacy, *privacy preservation becomes a goal that rationally interests other individuals*.
- Hence, privacy preservation becomes more attractive and therefore more sustainable.

Formalization of co-utility

Let Π be a game with self-interested, rational players $P^1, \cdots, P^N$, with $N > 1$. Game Π is said to be *co-utile* with respect to the vector $U = (u_1, \cdots, u_N)$ of utility functions if there exist at least two players P^i and P^j, having strategies s^i and s^j, respectively, such that: i) s^i involves P^i expecting co-operation from P^j; ii) s^j involves P^j co-operating with P^i; iii) (s^i, s^j) is an equilibrium for P^i and P^j in terms of u_i and u_j, respectively. In other words, there is co-utility between P^i and P^j, for some $1 \leq i, j \leq N$ with $i \neq j$, if the best strategy for P^i involves expecting co-operation from P^j and the best strategy for P^j is to co-operate.

Formalization of co-utility (II)

- For $\delta \in [0, 1]$, Π is said to be δ-co-utile with respect to U if the probability of it being co-utile is at least δ.
- A protocol is said to be co-utile if it has an underlying co-utile game.
- If the utility functions consider only privacy (resp. security, functionality), co-utility becomes co-privacy (resp. co-security, co-functionality).

Nash co-utility

- If players' strategies are pure and the equilibrium in co-utility is a Nash equilibrium, we have **Nash co-utility** (resp. **Nash δ-co-utility**).
- If mixed strategies are allowed and the equilibrium is Nash, then we have **mixed Nash co-utility** (resp. **mixed Nash δ-co-utility**).

Correlated co-utility

- The outcome of independent rational behavior by users, provided by Nash equilibria, can be inferior to a centrally designed outcome, that is, a correlated equilibrium.
- This is illustrated by the famous Prisoners' Dilemma (and by the current financial crisis!).
- Using correlated equilibria yields **correlated co-utility** (resp. **correlated δ-co-utility**).

Stackelberg co-utility

- Stackelberg equilibria are used when one player can impose strategies to the rest, as in attack-defense games (the attacker imposes her strategy).
- Such equilibria yield **Stackelberg co-utility** (resp. **Stackelberg δ-co-utility**).

Applications: the private information retrieval game

- Private information retrieval (PIR) is a game between a user and a database: the user wants to retrieve an item without the database knowing which.
- Most PIR protocols are ill-suited for PIR from a search engine or large database, because:
  - Their complexity is linear in the database size;
  - They (unrealistically) assume active cooperation by the database in the PIR protocol.

Pragmatic PIR relaxations

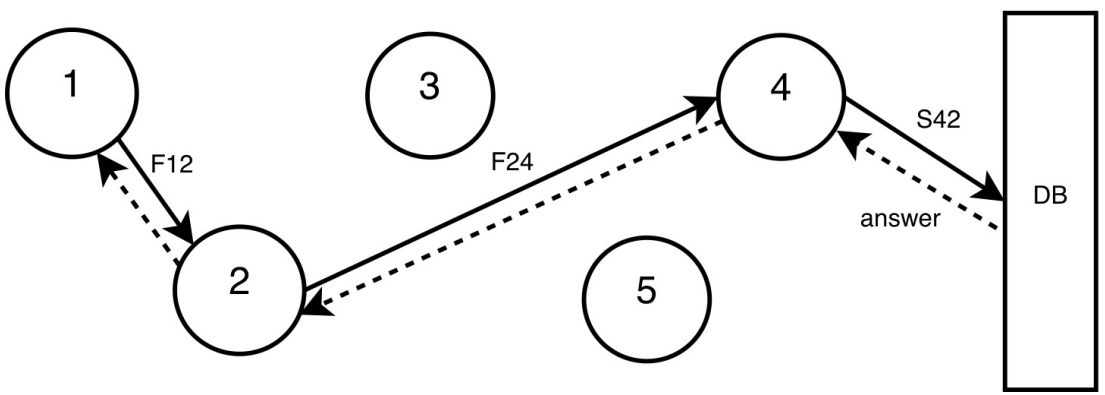
Standalone A program running locally in the user's computer

- either keeps submitting fake queries to cover the user's real ones (TrackMeNot)
- or masks the real query keywords with additional fake keywords (GooPIR).

P2P A user gets her queries submitted by other users in the P2P community; thus, the database still learns which item is being retrieved, but it cannot obtain the real query histories of users. This is user-private information retrieval (UPIR) or anonymous keyword search.

The P2P anonymous keyword search game

- Consider a system with N peers P^1 to P^N, who are interested in querying a database DB while keeping their interests (query profile) private.
- If P^i originates a query for submission to DB, she may submit it directly to DB or forward it to some P^j $(j \neq i)$.
- P^j may submit the query on P^i's behalf and return the results to P^i, or forward the query to some other P^k, with $k \notin \{i, j\}$. behalf and return the query results.



Privacy in P2P anonymous keyword search

- P^i's query interests stay private vs P^j (who does not know whether P^i is the query originator or a mere forwarder).
- The relevant privacy is of peers vs DB.
- Let $Y^i(t)$ be the set of queries submitted by P^i to DB up to time t.
- A plausible privacy utility function u_i for P^i is the Shannon entropy $H(Y^i(t))$: the higher this entropy, the flatter the histogram of frequencies of $Y^i(t)$ and the more ignorant DB stays about P^i's query interests.

Co-privacy in P2P anonymous keyword search

IF

- P^i decides to forward a query to P^j in order to avoid a submission that would "unflatten" her profile $Y^i(t)$;
- AND P^j decides to submit P^i's query because doing so makes his $Y^j(t)$ flatter;

THEN there is Nash co-utility and more specifically Nash co-privacy between P^i and P^j.

Types of privacy in social networks

- Content privacy. The information published by a user clearly affects her privacy, depending on how confidential is the information and to whom is it released.
- Relationship privacy. In some SNs, a user can specify how much it trusts other users and can establish several types of relationships with other users; knowing who is trusted by whom and to what extent discloses a lot about the users' thoughts and feelings.

A risk score for content disclosure in social networks

Liu-Terzi privacy risk score

Let the information attributes published by the users in an SN be labeled from 1 to n. Then the privacy score risk of user j is

$$PR(j) = \sum_{i=1}^{n} \beta_i \times V(i,j)$$

where $V(i,j)$ is the visibility of user j's attribute i (0 secret, 1 public) and β_i the attribute's sensitivity.

A utility function in the content disclosure game

The utility for user j is

$$PRF(j) = \frac{\sum_{j'=1, j' \neq j}^{N} \sum_{i=1}^{n} \beta_i V(i, j')}{1 + PR(j)}$$

$$= \frac{\sum_{j'=1, j' \neq j}^{N} \sum_{i=1}^{n} \beta_i V(i, j')}{1 + \sum_{i=1}^{n} \beta_i V(i, j)}$$

If independent strategies are used, we get the Prisoner's dilemma and the Nash equilibrium is for no user to reveal any content $\Longrightarrow$ the SN is cancelled.

The prisoner's dilemma in the content disclosure game

Users: u_1, u_2, each with one attribute.
Strategies: hide attribute (H), publish attribute (P).
Utility matrix:

User 2 / User 1	H	P
H	0, 0	1, 0
P	0, 1	1/2, 1/2

$\Longrightarrow$ (H, H) is a dominant strategy and a Nash equilibrium.

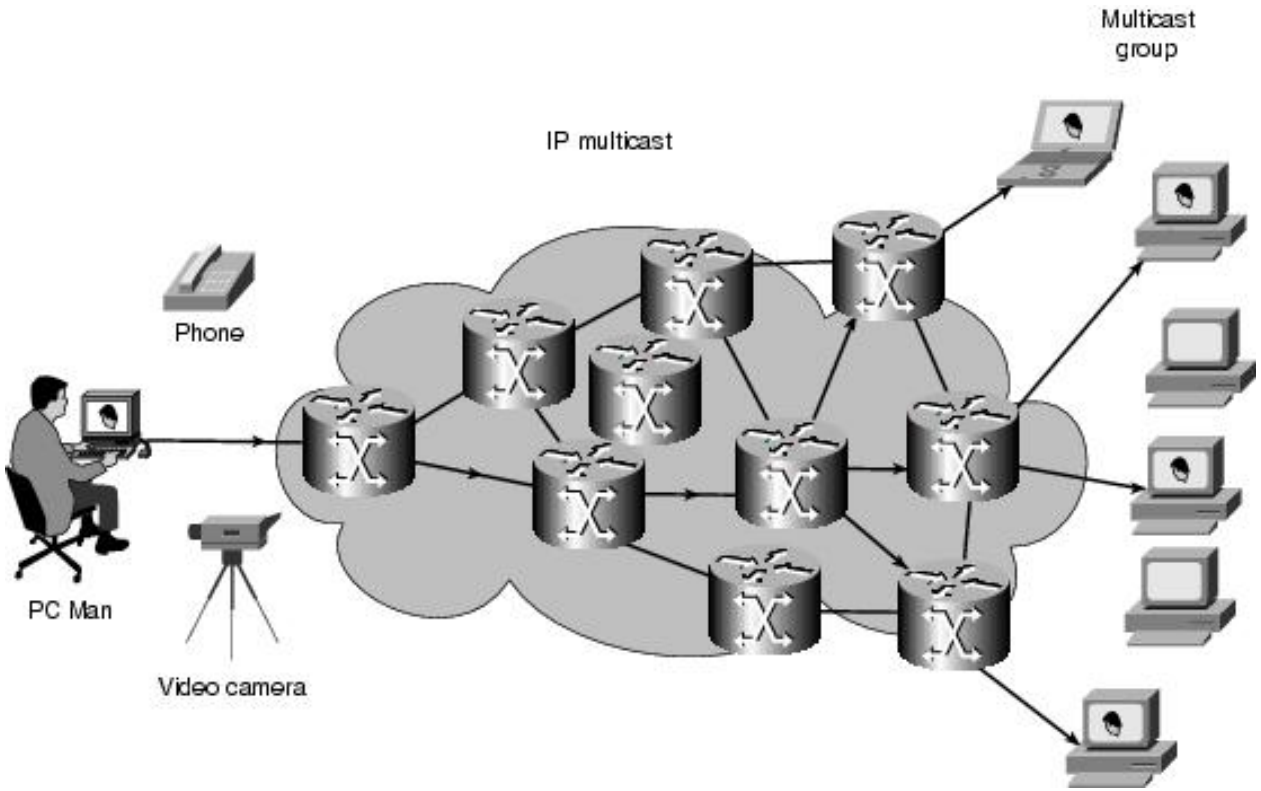
Correlated equilibria for content disclosure in SNs

- The outcome of independent rational behavior by users, provided by Nash equilibria and dominant strategies, can be inferior to a centrally designed outcome. This is clear above: the strategy (P, P) would give more utility than (H, H) to *both* users.
- However, usually no trusted third-party accepted by all users is available to enforce correlated strategies; in that situation, the problem is how User 1 (resp. User 2) can guess whether User 2 (resp. User 1) will choose P.

Correlated co-utility for content disclosure in SNs

- Correlated equilibria (based on tit-for-tat or reputation) yield more satisfactory, co-utile solutions.
- If SNs had a cryptographic infrastructure, cryptographic protocols could also be used:
  - Bitwise fair exchange of secrets;
  - Secure multiparty computation, etc.

Multicast



Fingerprinted multicast

- Each receiver must receive a fingerprinted copy, which is infeasible using standard multicast.
- Two alternatives:
  - Unicast transmission of a fingerprinted copy to each receiver (bandwidth-inefficient!).
  - Co-utile protocol in a P2P network for peers to co-operate in disseminating and anonymously fingerprinting copies.

Co-utility in fingerprinted multicast

- Utility function for peers includes reward and punishment
- After engaging in anonymous fingerprinting with P^{i+1}, P^i gets a reward $r_{i,i+1}$ from P^{i+1}, who discounts $r_{i,i+1}$ from her payoff d_{i+1}
- If P^i does not make any transfer to any other party traceable by the content source, P^i incurs an expected negative payoff $-p_i$ (cost of being accused and identified).
- With the above utility, there is co-utility between P^i and P^{i+1} for $i = 0$ to $N - 1$: indeed P^{i+1} obtains the content without losing her anonymity and P^i maximizes her utility by correctly fingerprinting it and making the transfer traceable to the content source.

Digital oblivion and co-utility

- Digital oblivion can be implemented by watermarking an expiration date into the content and fingerprinting the successive transfers to trace who redistributes/uses the content after the expiration date.
- The previous multicast fingerprinting scheme can be applied with slight modifications.

Cryptographic primitives for gradual privacy/utility

- To attain co-utility, it must be possible to trade off privacy against functionality and security.
- In correlated co-privacy (*e.g.* content disclosure in SNs), gradual privacy is required.
- We sketch some cryptographic primitives for gradual privacy that we have recently developed.

Asymmetric group key agreement (EUROCRYPT 2009)

- A (symmetric) group key agreement protocol (GKA) allows a set of users to establish a common secret key via open networks.
- We defined in Wu *et al.* (2009) asymmetric group key agreement (ASGKA).
- In ASGKA, only a shared encryption key is negotiated instead of a common secret key.
- This common encryption key is public and corresponds to different decryption keys, each of which is only computable by a group member.
- ASGKA allows encrypting to a temporary group.
- Applications of ASGKA include: broadcast to *ad hoc* groups, file sharing, secure group chat, group purchase of encrypted content with identity privacy, etc.

Contributory Broadcast Encryption (ASIACRYPT 2011)

- Broadcast encryption (BE) allows a sender to securely broadcast to any subset of members, but requires a trusted party to distribute decryption keys.
- ASGKA allows a group of members to negotiate a common encryption key, but a sender cannot exclude any particular member from decrypting the ciphertexts.
- Contributory broadcast encryption (CBE), presented in Wu *et al.* (2011), bridges BE and ASGKA.
- In CBE, a group of members negotiate a common public encryption key, while each member holds a decryption key, and a sender seeing the public group encryption key can limit the decryption to a subset of members of his choice.

Hierarchical Attribute-Based Encryption

- Attribute-based encryption (ABE) allows encrypting to uncertain decryptors by means of an access policy specifying the attributes that the intended decryptors should possess.
- Key management in ABE becomes difficult when there is a large number of users and *a priori* detailed access policies are not always feasible in practice.
- Hierarchical attribute-based encryption (HABE) is more versatile:
  - Attributes are organized in a hierarchy;
  - Users with higher-level attributes can delegate access to users with lower-level attributes;
  - Detailed *a priori* access policies are no longer required, because they can be refined through delegation.

Conclusions

- The novel concept of co-utility and its special case co-privacy have been introduced.
- Co-privacy makes privacy an attractive feature in P2P scenarios and co-utility can also improve security and functionality.
- Anonymous keyword search, content disclosure in SNs and fingerprinted multicast have been shown to be solvable with coprivate protocols.

Open issues

- Developing the theory of co-utility (mixed co-utility, Stackelberg co-utility, etc.);
- Developing new crypto protocols to implement the privacy graduality required by coprivacy (anonymous *ad hoc* broadcast encryption, new signatures, multiparty computation, etc.)

Related papers: theory of co-privacy/co-utility

- J. Domingo-Ferrer (2011) "Coprivacy: an introduction to the theory and applications of co-operative privacy", *SORT-Statistics and Operations Research Transactions*, vol. 35, special issue on privacy in statistical databases, pp. 25-40.
- J. Domingo-Ferrer (2010) "Coprivacy: towards a theory of sustainable privacy", in *Privacy in Statistical Databases-PSD 2010*, LNCS 6344, Springer, pp. 258-268.

Related papers: applications of co-privacy/co-utility

- J. Domingo-Ferrer and Ú. González-Nicolás (2012) "Rational behavior in peer-to-peer profile obfuscation for anonymous keyword search", *Information Sciences* 185(1):192-204.
- J. Domingo-Ferrer and Ú. González-Nicolás (2012b) "Rational behavior in peer-to-peer profile obfuscation for anonymous keyword search: the multi-hop scenario", *Information Sciences* (to appear).
- J. Domingo-Ferrer (2010b) "Rational privacy disclosure in social networks", in *MDAI 2010-Modeling Decisions in Artificial Intelligence*, LNCS 6408, Springer, pp. 255-265.
- J. Domingo-Ferrer (2011b) "Rational enforcement of digital oblivion", in *PAIS 2011-4th Intl. Workshop on Privacy and Anonymity in the Information Society*, ACM Digital Library.
- J. Domingo-Ferrer and D. Megías (2012) "Distributed multicast of fingerprinted content based on a rational peer-to-peer community", manuscript.

Related papers: crypto papers on gradual privacy

- Q. Wu, Y. Mu, W. Susilo, B. Qin and J. Domingo-Ferrer (2009) "Asymmetric group key agreement", in *EUROCRYPT 2009*, LNCS 5479, Springer, pp. 153-170.
- Q. Wu, B. Qin, L. Zhang, J. Domingo-Ferrer and O. Farràs (2011) "Bridging broadcast encryption and group key agreement", in *ASIACRYPT 2011*, LNCS 7073, Springer, pp. 143-160.
- H. Deng, Q. Wu, B. Qin, J. Domingo-Ferrer, L. Wang and H. Zhang (2012) "Hierarchical attribute-based encryption", manuscript.