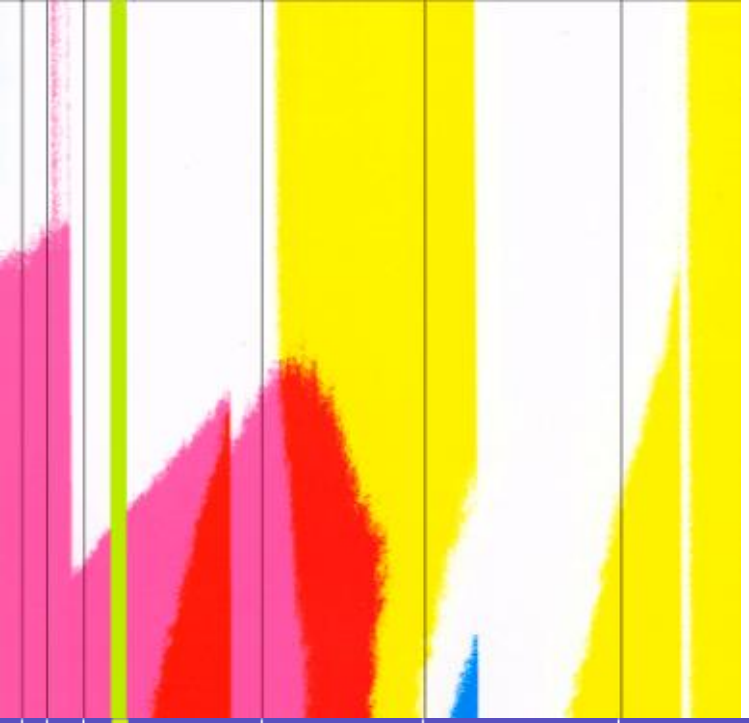

Privacy rights: new challenges, new approaches

Daniel Le Métayer

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA
RHÔNE-ALPES

Plan

1. Privacy as a fundamental right
2. New challenges
3. New approaches

Privacy: what are we talking about ?

- S. D. Warren – L. D. Brandeis (1890): Right to privacy as “right to be let alone”
- A. Westin (1967): Right for individuals to determine when how and to what extent information about them is communicated to others. Includes anonymity in public places: freedom from identification and surveillance
- P. Agre (1998): “Control over personal information is control over an aspect of the identity one projects to the world, and the right to privacy is the freedom from unreasonable constraints on the construction of one’s identity”

Fundamental values protected by privacy

- Instrumental to the protection of fundamental rights: liberty (opinions), equality (non discrimination)
- Benefits for the individual: self-realization, autonomy, dignity, etc.
- Benefits for society: secure the conditions for citizen participation in deliberative democracy, contribute to prevent the “tyranny of the majority”, protect dissent opinions, pressure to conform to dominant norms, etc.

Regulatory policies

- International instruments
- European instruments
- National instruments
- Guidelines, recommendations, codes of practice (by business sector).

International instruments: United Nations

- Universal Declaration of Human Rights (1948): “No one shall be subjected to arbitrary interference with his privacy, home or correspondence, nor to attacks upon his honor or reputation. Everyone has the right to the protection of the law against such interference or attacks”.
- Guidelines Concerning Computerized Data Files (1990): lawfulness, fairness, accuracy, purpose specification, interested person access, non discrimination, security, supervision and sanction, transborder data flows.

European instruments

Council of Europe:

- European Convention for the Protection of Human Rights and Fundamental Freedoms (1950): “everyone has the right to respect for his private and family life, his home and correspondence “
- European Court of Human Rights

European Union:

- Directive 95/46/EC on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data
- Directive 2002/58/EC (amended in 2009) concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector

European Directive 95/46/EC
Rights of the subject

- Unambiguous consent (with derogations: contractual or legal obligations, vital interest of the subject, etc.)
- Right of access: confirmation, communication of data, logic of the processing, rectification, notification to recipients
- Right to object: at any time, on request in case of direct marketing, “on compelling legitimate grounds” in general
- No decision producing legal effects based solely on automated processing of personal data (performance, creditworthiness, conduct, etc.)

European Directive 95/46/EC
Further protections

- Liability of the controller (person who determines the purposes and means of the processing of personal data)
- The controller is in charge (inter alia) of ensuring the confidentiality of personal data and the security of processing
- Limitations on the transfer of personal data to third countries (but as few as possible within EU)
- Independent privacy agencies with great powers (authorization, notification, control, injunction, arbitration, sanction, etc.). Examples: CNIL (France), Garante (Italy), BFD (Germany), etc.

National instruments: USA

- Comprehensive legislation for federal government agencies (Privacy Act, 1982)
- "Omnibus" legislative solutions for the private sector (FSPA: Financial Services Privacy Act, ECPA: Electronic Communication Privacy Act, HIPAA: Health Insurance Portability and Accountability Act, COPPA: Child Online Privacy Protection Act, etc.)
- "Safe harbour" agreement between the USA and Europe for the flow of personal data from EU to US-based companies abiding by a set of "fair information" principles

Limits to privacy

- Universal notion but levels and forms of privacy concerns vary a lot: cultural, historical, technological, personal factors
- Conflicting values: freedom of speech, right to be informed, public security, etc.
- Role of the state (USA: regulation only when market has failed, Europe: the state should protect individuals)
- New technologies, new practices, …

New technologies : new challenges to privacy

- New ways to collect personal data: on-line interactions, sensors, biometric devices, RFID tags, mobile devices, cameras, smart cards, GPS, GSM (either openly or without user's knowledge)
- New ways to exploit personal data: data mining, knowledge inference, behavioral modeling (possibly beyond the subjects' own knowledge)
- New incentives for users: economic incentives (free services, loyalty cards, etc.), location-based services, personalization, new facilities (on-line reservation, e-commerce, e-ticketing, etc.), enhanced security, etc.

New challenges : new answers

- Information flow is no longer the exception, it is the norm in the information society
- More and more difficult for subjects to effectively exercise their rights (consent, access, objection, deletion, etc.)
- New combinations of legal and technical instruments are required
- One way forward: privacy by design
  - Privacy by design in the current draft for the new European Data Protection Regulation (to replace the European Directive 95/46/EC)
  - Privacy by design from the technical point of view
  - Application to a location-based system

Privacy by design

Ann Cavoukian, Information and Privacy Commissioner of Ontario (November 2008):

“The purpose of privacy by design is to give due consideration to privacy needs prior to the development of new initiatives – in other words, to consider the impact of a system or process on individuals’ privacy and to do this throughout the systems lifecycle, thus ensuring that appropriate controls are implemented and maintained. ”

Europe: Working Party 29

December 2009:

- "The principle of "Privacy by Design" should be introduced in the new framework: privacy and data protection should be integrated into the design of Information and Communication Technologies. …
- This principle of "Privacy by Design" should not only be binding for data controllers, but also for technology designers and producers. On top of that, as the need arises, regulations for specific technological contexts should be adopted which require embedding data protection and privacy principles into such contexts."

Draft Regulation released by the EC (January 2012)

Recital 61:

The protection of the rights and freedoms of data subjects with regard to the processing of personal data require that appropriate technical and organisational measures are taken, both at the time of the design of the processing and at the time of the processing itself, to ensure that the requirements of this Regulation are met. In order to ensure and demonstrate compliance with this Regulation, the controller should adopt internal policies and implement appropriate measures, which meet in particular the principles of data protection by design and data protection by default.

Draft Regulation released by the EC (January 2012)

Article 23 (2):

The controller shall implement mechanisms for ensuring that, by default, only those personal data are processed which are necessary for each specific purpose of the processing and are especially not collected or retained beyond the minimum necessary for those purposes both in terms of the amount of the data and the time of their storage. In particular, those mechanisms shall ensure that by default personal data are not made accessible to an indefinite number of individuals.

Draft Regulation released by the EC (January 2012)

Article 23 (3):

The Commission shall be empowered to adopt delegated acts in accordance with Article 86 for the purpose of specifying any further criteria and requirements for appropriate measures and mechanisms referred to in paragraph 1 and 2, in particular for data protection by design requirements applicable across sectors, products and services.

The Commission may lay down technical standards for the requirements laid down in paragraph 1 and 2.

Case study: Pay as you Drive

Impact of the minimisation principle on the architecture of the system

Context

More and more initiatives around the world:

- Existing systems: California, London, Stockholm, Singapur, etc.
- Projects: New York, Boston, Netherlands, etc.

Decision of the European Commission (october 2009): standardization of road fee systems (Service Européen de Télépéage)

Expected benefits:

- Reduction of traffic jams
- Reduction of pollution
- Fairness

First option (centralized)

- On Board Equipment (OBE)
  - GPS
  - GSM
  - On board computer : sends all location data and vehicle identification to the server
- Operator
  - Computes the fee due for each car
  - Spot-checks the cars to detect misbehaviours or failures of the OBEs

First option (centralized)

Secure solution for the operator but

- The operator knows all the whereabouts of all the vehicles $\Rightarrow$
- Highly privacy intrusive

Second option (Vpriv)

- On Board Equipment
  - Commits to a fixed set of anonymous tags
  - Sends the location data to the server with anonymous tags
  - Adds the fees corresponding to its own tags and returns the sum to the server
- Operator
  - Computes the fee due for each location data received
  - Returns to each car all the individual fees (end of each quarter)
  - Spot-checks the cars to detect misbehaviors or failures of the OBEs
  - Conducts the verification protocol to check the sum returned by the OBEs (dedicated protocol for secure multi-party computation)

Second option (Vpriv)

Better solution for the driver but

- Requires anonymous communications
- Risks of de-anonymization
- Complexity and cost

Third option (Secure OBE)

- On Board Equipment
  - Secure component
  - Performs all the computations of the fees
  - Sends the fee to the operator at the end of each quarter
- Operator
  - Spot-checks the cars to detect misbehaviors or failures of the OBEs (two-way communications)

Third option (Secure OBE)

Excellent solution w.r.t. data minimization but

- Requires more expensive OBEs
- Need to update the fee calculation software securely

Forth option (Commitments)

- On Board Equipment
  - Sends vehicle identification and hashes of the location data to the server
  - Performs the computations of the fees
  - Sends the fee to the operator at the end of each quarter
  - Discloses partial sums in case of spot checks
- Operator
  - Spot-checks the cars to detect misbehaviors or failures of the OBEs
  - Conducts the verification protocol

Communications

VH $\rightarrow$ OP: $\langle id, j, h(\theta_{j,1}), \ldots, h(\theta_{j,144})\rangle$ every day

VH: $t = \Sigma_{j,i} F(\theta_{j,i})$

VH $\rightarrow$ OP: $\langle id, t\rangle$ every quarter

id: vehicle identifier

j: day

$\theta_{j,i}$: trajectory of the vehicle for day j (e.g. every 10 minutes)

t: fee due for the quarter

Verification

The operator knows:

- $h_{j,i}$
- t

He wants to check:

$t = \Sigma_{j,i} F(\theta_{j,i})$

Spot checks

OP: p_k

OP $\rightarrow$ VH: trajectory corresponding to p_k ?

VH $\rightarrow$ OP: $\theta_{j,i}$

OP: $h_{j,i} =_? h(\theta_{j,i})$ and $p_k \in_? \theta_{j,i}$

spot checks

after receipt of $h_{j,i}$

Spot checks can be done without interaction with the vehicles

Verification protocol (1/2)

OP $\rightarrow$ VH: breakdown of t over the 3 months ?

VH $\rightarrow$ OP: $< m_1 , m_2, m_3 >$

OP: $t =_? m_1 + m_2 + m_3$

OP $\rightarrow$ VH: breakdown of the sum for the second month ?

VH $\rightarrow$ OP: $< s_1 , s_2, s_3 , s_4 >$

OP: $m_2 =_? s_1 + s_2 + s_3 + s_4$

OP $\rightarrow$ VH: breakdown of the sum for the forth week ?

VH $\rightarrow$ OP: $< j_1 , j_2, j_3 , j_4 , j_5 , j_6, j_7 >$

OP: $s_4 =_? j_1 + j_2 + j_3 + j_4 + j_5 + j_6 + j_7$

Verification protocol (2/2)

OP $\rightarrow$ VH: breakdown of the sum for the first day ?

VH $\rightarrow$ OP: $< v_1, \ldots, v_{144} >$

OP: $j_1 =_? v_1 + \ldots + v_{144}$

OP $\rightarrow$ VH: $\theta_{50,36}$?

VH $\rightarrow$ OP: $\theta_{50,36}$

OP: $h_{50,36} =_? h(\theta_{50,36})$ $v_{36} =_? F(\theta_{50,36})$

j_1 : 50^e day of the quarter

Forth option (Commitments)

Flexible but

- Interactive verification protocol
- Non minimal disclosure of data during spot checks

Enhancements:

- Commitment trees (de Jonge – Jacobs)
- Homomorphic commitments (Balash et. al. : PrETP)

Beyond specific technical choices

- Current situation :
  - A range of techniques are already available (PETs): anonymization, commitments, secure multiparty computation, homomorphic encryption, sanitization (adding noise, clustering, perturbation, aggregation, sampling, etc.)
  - Had hoc solution for each problem: difficult to build on past experience, difficult to compare solutions, not cost effective
- What we need:
  - A more systematic approach to privacy by design
  - Going from art to an industrial process

Towards a systematic approach to privacy by design

Definition of the context:

- Service to be performed
- Actors involved
- Requirements of each actor

Objective: exploration of the design space

- Find architectures which can both deliver the service and meet the constraints
- Check if an architecture meets the constraints and can deliver the service

Formal model for data minimization

- Service: set of equations
- Requirements of the parties: constraints on variables (actors who can collect, spot check, control, … each variable)
- Exploration of the design space: inference system

Notation:

$\rho(X) = A$ for "A controls the computation of the equation defining X"

Illustration : Pay as you drive

- Actors: VH, OP and Env
- Service:
  - $T = M_1 + M_2 + M_3$ — Fee for a quarter
  - $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
  - $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
  - $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
  - $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Illustration : Pay as you drive

Operations:

- $Receive_{A,B}$ (X,V) — A receives the value V of variable X from B
- $Get_{A,B}$ (X,V) — A spot-checks the variable X from B
- $Commit_{A,B}$ (X,V) — A receives the commitment V for X from B
- $Check_{A,B}$ (X,V,D) — A discovers and checks the value V of X
- $Compute_{A}$ (X,V) — A computes the value V of X

Formel system

Inference system :

- $C \vdash_A X$ in architecture C, A can detect an error in the computation of X
- $C = \{(R_i, D_i, G_i)\}$
  - R_i : set of variables which can be received by A in a run
  - D_i : set of variables which can be committed and discovered by A in a run
  - G_i : set of variables which can be spot-checked by A in a run

C : architecture of the system (set of operations available to A to perform his verifications)

Illustration : Pay as you drive

Equations :

- $T = M_1 + M_2 + M_3$ — Fee for a quarter
- $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
- $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
- $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
- $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Requirements for the first option:

- $\rho(T) = \rho(M_i) = \rho(J_{i,j}) = \rho(H_{i,,j,k}) = OP$
- $\forall (R,D,G) \in C, R \subseteq \{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}$
- $\forall (R,D,G) \in C, G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge Card(G) \leq 1$
- $C \vdash_{OP} T$

Inference system

Rule 1:

$$\forall i \in \{1,\ldots,n\},\ U_i \vdash_A Y_i \qquad X=F(Y_1,\ldots,Y_n) \in Eq$$

$$\{(R \cup R', D \cup D', G \cup G') \mid (R,D,G) \in \cup_{i \in \{1,\ldots,n\}} U_i \} \vdash_A X$$

$$R' \cup D' \cup G' = \{X, Y_1, \ldots, Y_n\}$$

Intuition:

A must be able to obtain the values of all the variables one way or another and to check the correctness of any of them

Inference system

Rule 2:

$$\frac{\forall i \in \{1,\ldots,n\},\ U_i \vdash_A Y_i \qquad X=F(Y_1,\ldots,Y_n) \in Eq \qquad \rho(X) = A}{\cup_{i \in \{1,\ldots,n\}} U_i \vdash_A X}$$

Intuition:

Because A controls the computation of X, he must just be able to check any of the input variables Y

Inference system

Rule 3:

$$\frac{U \vdash_A X \qquad U \leq U'}{U' \vdash_A X}$$

with

$U \leq U' \Leftrightarrow \forall (R,D,G) \in U,\ \exists (R',D',G') \in U',\ R \subseteq R', D \subseteq D', G \subseteq G'$

Intuition:

Any verification remains possible in an enriched context

Semantics

- Semantics over distributed traces: effect of each operation on the knowledge set of each actor
- Assumptions on traces: properties of cryptographic operations, notion of control, threat model (tampering with variables)
- Proof of correctness of the inference system :

  $C \vdash_A X$ and $\mathrm{Wrong}(X, \sigma) \Rightarrow$

  A can use the operations allowed in C to extend the trace σ into a trace σ' which brings to A the proof that X is not correct

First option (centralized)

Equations :

- $T = M_1 + M_2 + M_3$ — Fee for a quarter
- $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
- $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
- $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
- $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Requirements :

- $\rho(T) = \rho(M_i) = \rho(J_{i,j}) = \rho(H_{i,,j,k}) = OP$
- $\forall (R,D,G) \in C, R \subseteq \{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}$
- $\forall (R,D,G) \in C, G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge Card(G) \leq 1$
- $C \vdash_{OP} T$

First option (centralized)

$$C = \{(P, \varnothing, \{\theta_{i,,j,k}\}) \mid i \in Di, j \in Dj, k \in Dk\} \vdash_{OP} T$$

with

$$P = \{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}$$

Architecture C meets the requirements:

- $\forall (R,D,G) \in C, R \subseteq \{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}$
- $\forall (R,D,G) \in C, G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge Card(G) \leq 1$

Other option (decentralized)

Equations :

- $T = M_1 + M_2 + M_3$ — Fee for a quarter
- $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
- $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
- $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
- $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Requirements :

- $\rho(T) = \rho(M_i) = \rho(J_{i,j}) = \rho(H_{i,,j,k}) = OP$
- $\forall (R,D,G) \in C, R \subseteq \{T\} \wedge D = \varnothing \wedge$
  $G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge Card(G) \leq 1$
- $C \vdash_{OP} T$

Other option (decentralized)

Impossible to prove

$$C \vdash_{OP} T$$

with an architecture C meeting the requirements

Intuition:

OP can control the computation but not the input data ($P_{i,,j,k} = \theta_{i,,j,k}$)

Third option (Secure OBE)

Equations :

- $T = M_1 + M_2 + M_3$ — Fee for a quarter
- $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
- $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
- $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
- $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Requirements :

- $\rho(T) = \rho(M_i) = \rho(J_{i,j}) = \rho(H_{i,,j,k}) = OP$
- $\forall (R,D,G) \in C, R \subseteq \{T\} \wedge D = \varnothing \wedge$
  $G \subseteq (\{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \cup \{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}) \wedge Card(G) \leq 2$
- $C \vdash_{OP} T$

Third option (Secure OBE)

$$C = \{(\{T\}, \varnothing, \{\theta_{i,,j,k}, P_{i,,j,k}\}) \mid i \in Di, j \in Dj, k \in Dk\} \vdash_{OP} T$$

Architecture C meets the requirements :

$$\forall (R,D,G) \in C, R \subseteq \{T\} \wedge D = \varnothing \wedge$$
$$G \subseteq (\{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \cup$$
$$\{P_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\}) \wedge \text{Card}(G) \leq 2$$

Forth option (Commitments)

Equations :

- $T = M_1 + M_2 + M_3$ — Fee for a quarter
- $M_i = J_{i,1} + \ldots + J_{i,31}$ — Fee for a month
- $J_{i,j} = H_{i,j,,1} + \ldots + H_{i,j,144}$ — Fee for a day
- $H_{i,,j,k} = F(P_{i,,j,k})$ — Fee for a 10 minutes period
- $P_{i,,j,k} = \theta_{i,,j,k}$ — $\theta_{i,,j,k}$: effective trajectory

Requirements :

- $\rho(T) = \rho(M_i) = \rho(J_{i,j}) = \rho(H_{i,,j,k}) = VH$
- $\forall (R,D,G) \in C, R \subseteq \{T\} \wedge$
  $G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge Card(G) \leq 1$
- $C \vdash_{OP} T$

Forth option (Commitments)

$$C = \{(\{T\}, D_{i,j,k}, \{\theta_{i,,j,k}\}) \mid i \in Di, j \in Dj, k \in Dk\} \vdash_{OP} T$$

with $D_{i,,j,k} = = \{P_{i,,j,k}\} \cup \{H_{i,,j,\_}\} \cup \{J_{i,\_}\} \cup \{M_i\}$

The set of ancestors and collaterals of $\theta_{i,,j,k}$ in the computation tree of T

Architecture C meets the requirements:

$$\forall (R,D,G) \in C, R \subseteq \{T\} \wedge$$
$$G \subseteq \{\theta_{i,,j,k} \mid i \in Di, j \in Dj, k \in Dk\} \wedge \text{Card}(G) \leq 1$$

Benefits of the formal approach

- Precise definitions of assumptions and requirements
- Detection of inconsistencies
- Systematic exploration of the design space

Extensions and further work

- Data minimisation principle:
  - Higher level requirement language: data inference
  - Minimality analysis of the set of equations defining the service
  - Transformation of the set of equations
  - Probabilistic framework
- Informed consent of the subject
- Rights of the subject
- Transparency
- etc.

Privacy by design - Conclusion

- Consensus on the fact that the privacy by design approach should be promoted, supported by legal instruments and more widely adopted
- Consensus on general principles (data minimization, informed consent, transparency, etc.)
- A range of techniques are already available (anonymization, commitments, secure multiparty computation, homomorphic encryption, etc.)
- What is badly needed is a systematic approach and tools to support it
- Privacy is a complex issue with potentially conflicting requirements $\Rightarrow$ Formal methods can be play an instrumental role in this context

Bibliography

J. Balasch, A. Rial, C. Troncoso, C. Geuens, B. Preneel, I. Verbauwhede, PrETP: Privacy-Preserving Electronic Toll Pricing, Proc. 19th USENIX Security Symposium, 2010.

A. Cavoukian, Privacy by design, Report of the Information & Privacy Commissioner Ontario, Canada

F. Garcia, B. Jacobs, Privacy-friendly energy metering via homomorphic encryption, Proc. 6th Workshop on Security and Trust Management, Springer Verlag LNCS, 2010.

W. De Jonge, B. Jacobs, Privacy-friendly electronic traffic pricing via commits, Proc. Workshop of Formal Aspects of Security and Trust, Springer Verlag, LNCS 5491, 2009.

D. Le Métayer, Privacy by design: a matter of choice, in Data Protection in a Profiled World, S. Gutwirth, Y Poullet, P. De Hert, Springer Verlag, 2010

D. Le Métayer, A formal privacy management framework, Proc. Workshop of Formal Aspects of Security and Trust, Springer Verlag, LNCS 5491, 2009.

A. Popa, H. Balakrishnan, A. Blumberg, Vpriv: protecting privacy in location-based vehicular services, Proc. 18th USENIX Security Symposium, 2009.