Privacy Technologies and Applications on Electronic Markets
Popular privacy policies

We automatically track certain information based upon your behavior on the Site. We use this information to do internal research on our users' demographics, interests, and behavior to better understand, protect and serve you and our community. 

(ebay.com)

We receive and store any information you enter on our Web site or give us in any other way. 

(a9.com)

We will retain in our files some personal information you have requested to remove [...] Therefore, you should not expect that all of your personal information will be completely removed from our databases in response to your requests. 

(ebay.com)

We employ other companies and individuals [...] They have access to personal information needed to perform their functions, but may not use it for other purposes. 

(amazon.com)

Google may share cookie information among its other services for the purpose of providing you a better experience. 

(gmail.google.com)
What is Privacy?

– Privacy in the context of Computer Science

• Data Security encompasses three aspects
  - Confidentiality (C)
  - Integrity (I)
  - Availability (A)

• Privacy is one aspect of confidentiality.

• Main challenge: Secure information instead of data.
Privacy and related topics

- Privacy
- Civil protection
- Data security
- IT security
- Reliability

[Pommerening 1994, 2004]
Types of Privacy

– Four types of Privacy
  
  • Physical Privacy
  • Mental Privacy
  • Decisional Privacy
  • Informational Privacy

– Alternative Privacy classifications exist.
Information Privacy

– Information privacy is a concern
  • in private life
  • in public life
  • in professional life

– Classical justification for information privacy
  • Personal information is a right (Warren and Brandeis in Harvard Law Review 1890, iv(5), 193-220
  • Control of the information has to be owned by the individual
Infringements of Information Privacy

– by governmental agencies
  • police, Dept. of Homeland Security, Secret Service, …
  • parafisci (insurances, churches, …)

– by commercial organizations
  • single enterprises
  • groups of enterprises (what about the data in case of M&A?)

– by individuals
Protecting Privacy

– legal measures
  • huge amount of privacy-related laws
  • “Pyramid of Norms”

– social measures
  • e.g. privacy sensitization, education

– organizational and economical measures
  • self-regulation, codes of conduct
  • market mechanisms

– technical measures
Actors involved in privacy protection

– Legislation organizations
– Ombudsman, Data Protection Officer
– Consumer protection organization
– Certification authorities
– the individual himself
Privacy as a right

- Human Right
- Fundamental Right
  - National constitutions
  - European Constitution
- Personal Right
- Right on informational self-determination
Privacy legislation

- International legislation
  - United Nations

- European legislation
  - Directly taking effect
  - Implemented in national laws

- Transnational legislation
  - e.g. Safe Harbor agreement

- National legislation
  - Federal laws
  - Federal States laws
  - Application-specific laws (lex specialis)

- Case law
Privacy legislation layers

– Content layer

– Service layer

– Transport layer
Privacy legislation in Germany (Selection)

- Bundesdatenschutzgesetz (BDSG), Landesdatenschutzgesetze
- Bundesstatistikgesetz (§11), Landesstatistikgesetze
- Hochschulstatistikgesetze
- Meldegesetz
- Verwaltungsverfahrensgesetz (§4, §5, §30, §84)
- Sozialgesetzbuch (§35: Sozialgeheimnis)
- Fernmeldeanlagengesetz
- Urheberrechtsgesetz
- Strafgesetzbuch (§202a, §263a, §269, §303a, §303b)
- Bundeskrebsregistergesetz, Landeskrebsregistergesetze
- Grundsätze ordnungsgemäßer Speicherbuchführung
- Datenschutz-Richtlinie der EU von 1995
- Signaturgesetz, Signaturverordnung, Begründung zur Verordnung zur digitalen Signatur
- Teledienstdatenschutzgesetz (TDDSG)
- Mediendienstestaatsvertrag (MDStV)
- Telekommunikationsgesetz (TKG)
- …
European privacy legislation (1)

– Legislative documents (selected documents)
  • Treaty on the European Union (TEU): Article F
  • European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR): Art. 8
  • EU Charter of Fundamental Rights of 7 December 2000
European privacy legislation (2)

- Legislative documents (selected documents)
  - Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
European privacy legislation (3)

– Case law

• Judgment of the European Court of Justice of 20 May 2003
  Joint Affairs C-465-00, C-138/01 and C-139/01

• Judgment of the Court in Case C-101/01 - Bodil Lindqvist

[http://europa.eu.int/comm/internal_market/privacy/law_en.htm]
OECD Guidelines
on the Protection of Privacy and Transborder Flows of Personal Data

– Basic Principles
  • Collection Limitation Principle
  • Purpose Specification Principle
  • Use Limitation Principle
  • Security Safeguards Principle
  • Openness Principle
  • Individual Participation Principle
  • Accountability Principle
Federal Trade Commission (US)

– Fair Information Practice Codes
  • Notice / Awareness
  • Choice / Consent
  • Access / Participation
  • Integrity / Security
  • Enforcement / Redress
    - Self-Regulation
    - Private Remedies
    - Government Enforcement

– Application to Information Collected From Children
Canadian Standards Association

- Model Code for the Protection of Personal Information, CAN/CSA-Q830-95
- “This model code describes the minimum requirements for the protection of personal information. Any applicable legislation must be considered in implementing these requirements.”
- “The objective of this Standard is to assist organizations in developing and implementing policies and practices to be used when managing personal information.”
- „Ten interrelated principles form the basis of the CSA Model Code for the Protection of Personal Information.“
Privacy Code Principles / CSA

- Accountability
- Identifying Purposes
- Consent
- Limiting Collection
- Limiting Use, Disclosure, and Retention
- Accuracy
- Safeguards
- Openness
- Individual Access
- Challenging
Main Threats to Privacy

- Intrusion and Theft
- Communication
- Forwarding, Sharing
- (Secretly) Linking of personal data
- Malpractice
- Usage of personal data with unintentional purpose
- Exceeded retention time
How can consumers protect themselves?

– Approach 1: Refuse online activities

– Approach 2: Minimize availability of personal data
  • Data economizing
  • Controlled transmission (e.g. Identity Management)

– Approach 3: Minimize feasibility of linking
  • Between different transactions
  • Between different online identities
  • Between online identity and offline identity

– Approach 4: Do not care about personal data
Privacy Enhancing Technologies

“Privacy-enhancing technologies are protocols, standards and tools that directly assist in protecting privacy, minimizing the collection of personally identifiable information, and when possible, eliminating the collection of personally identifiable information”

[cited from [Agrawal 2003], based on [Burkert 1997], [Clarke 2002], [Rotenberg 2001]]
PETs

– 2 categories of PETs: [Fischer-Hübner]
  • Products which provide consumer choice, such as P3P (Platform for Privacy Preference Project) from W3C
  • Products which protect user identity through
    - Anonymity
    - Pseudonymity
    - Unlinkability
    - Unobservability

– Enabling PETs

Organizational PETs

  • encryption
  transparency and trust
PITs

– PITs are the counterparts of PETs
  • Privacy Invasive Technologies
  • software based mechanisms that enable privacy infringements

– Examples: Mechanisms for …
  • data collection
  • data mining
  • data merging
  • …
  • enabling technologies (e.g. storage solutions)
Data collection (examples)

– Offline
  • Video cameras
  • Rebate cards (PayBack, HappyDigits, …)
  • Toll
  • RFID
  • Cell phones
  • Electronic Cash transactions
  • Call / connection data

– Online
  • Cookies
  • Referer
  • Webbugs
  • General HTTP transmission data
  • Connection data (ISP)
Re-identification (inferences)

An inference occurs when one is able to infer some data without directly accessing it.

- e.g. two data attributes may allow to infer a third one
- \( f(d_1, \ldots, d_n) \rightarrow d_{n+1} \quad j \neq k \Rightarrow d_j \neq d_k \)

Inferences

- Learned from existing database
- Deduced from external knowledge
Inferences – Classical example

– Sweeney, 2001:
  - sample of 54,805 voters in Cambridge
  - she was able to identify
    - 12% based on their birth dates
    - 29% based on birth date and gender
    - 69% based on birth date and 5-digit zip code
    - 97% based on birth date and 9-digit zip code
General inference mechanism

– Overlapping knowledge and External knowledge
Re-identification and k-anonymity

– Easier re-identification by:
  • More sources of information and consistent compatible ontologies
  • More overlapping knowledge
  • Differentiated attributes, low aggregation level

– k-anonymity [Sweeney, 2002a, 2002b]
  • “A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release.”
Identity and attribute disclosure

– Identity disclosure
  • By record linkage across databases
  • By rare or extreme attribute values

– Attribute disclosure
  • Value of sensitive attribute is disclosed
  • Does not require identity disclosure
Methods to prevent re-identification

- remove identifiers
  - remove explicit identifiers (name and address, SSN, …)
  - Remove implicit identifiers (“occupation = Mayor of New York”)
  - aggregation (geographical or otherwise)
  - top-coding (lumping attribute values into a single category)

- distort disclosure-inducing high-dimensional statistical characteristics:
  - cell suppression
  - data swapping
  - jittering (adding random noise)

- analysis servers (which disseminate analyses of data rather than data themselves)
Cookies

– History
  • 1994: introduced as Netscape’s proprietary extension to the HTTP
  • 1997: standardised as „HTTP State Management Mechanism“ by the IETF in RFC 2109

– Why cookies?
  • HTTP has no own mechanism for realising sessions
  • every request is treated separately
  • a priori: all the data has to be entered again for every single request
  • solution: cookies save a small amount of data
The cookie concept

- cookies save a small amount of data:
  - administrative information, such as:
    - server that set the cookie
    - server and path restrictions for the cookie visibility
    - date of expiry
  - content
    - name-value-pairs
    - limited in capacity
- the cookie is persistent on the user’s computer
  - for the current session only (session-cookies)
  - over multiple sessions (persistent cookies)
Re-identification by cookies

- re-identify users over multiple sessions / pages
- associate anonymous data with identifiable data

Identified clickstream

Page 1 \[\rightarrow\] Page 2 \[\rightarrow\] Page k \[\rightarrow\] Page n

Server sets cookie

User enters identity
Linkable and Linked Data

- Linking cookie data and user profiles
  - Cookies often store a unique identifier (e.g. session id, database key) that links to the complete profile in the database

- Examples
  - data stored directly in the cookie:

```
notissesel
766064:Video%20der%20Neujahrssprache%20von%20Bundeskanzler%20Gerhard%20
www.bundesregierung.de/
```

  - key stored in the cookie:

```
suid
ec0-baltwa4-4epj21-46169c3a373c dac9da7c17e57d5054fb939b05717f1db24a0
sueddeutsche.de/
```
Cookies and legislation

– EU privacy directive
  • Providers must provide users the possibility to refuse cookies.
  • They have to provide the user precise and understandable information about the aim of the cookies they want to set.

– Theoretically,
  • users are well-protected against cookie-related privacy issues.

– Practically,
  • this protection is not always respected.
Cookies and privacy issues

– Session-cookies
  • unproblematic
  • the user is not re-identifiable afterwards
  • equivalent to: session IDs in hidden input field or URI

– Persistent cookies
  • problematic
  • allow re-identification over long time spans
  • a warning in the browser or a notice in the terms and conditions are not sufficient
Webbugs

- Little (transparent) images included in a website
  - they can transmit information such as:
    - if and when a page is viewed
    - with scripting support:
      - operating system language
      - screen resolution
      - page dimensions
      - ...
  - they can only be identified in the source of the page
  - used in “normal” websites and e-mails
  - use case example: IVWBox (http://www.ivw.de, http://www.ivwonline.de)
Blocked webbugs in emails

Vielen Dank fuer Ihre Bestellung, Soeren Preibusch!

Moechten Sie den Stand Ihrer Bestellung einsehen?
Gehen Sie zu Mein Konto.
Automatically collected information at amazon.com

“Examples of the information we collect and analyze include the Internet protocol (IP) address used to connect your computer to the Internet; login; e-mail address; password; computer and connection information such as browser type and version, operating system, and platform; purchase history, which we sometimes aggregate with similar information from other customers to create features such as Purchase Circles, Top Sellers, and Just Like You; the full Uniform Resource Locator (URL) clickstream to, through, and from our Web site, including date and time; cookie number; products you viewed or searched for; zShops you visited; your Auction history; and the phone number you used to call our 800 number. During some visits we may use software tools such as JavaScript to measure and collect session information, including page response times, download errors, length of visits to certain pages, page interaction information (such as scrolling, clicks, and mouse-overs), and methods used to browse away from the page.”

[Amazon.com Privacy Notice 2003], own highlighting
IP-based identities

- **IP addresses**
  - (together with cookies and other credentials) form the basis for most notions of online identity

- **Applications**
  - using the IP address as an identifier
  - using the IP address as source of information
    - profiling of users that do not actively identify themselves.
    - a primitive form of localization
    - identifying dial-up connections, governmental agencies or institutions by their (dynamic or static IP addresses)
IP-based identities: Problems

– Identity owner
  • no way to turn it off completely
  • global identity

– Service provider
  • No 1:1 mapping of IP addresses and offline identities
  • Hard to trace
    - IP addresses *usually* map to real identities, but not always
IP-anonymity: mix networks (1)

– Implementation
  • Request: User → Proxy → Server
  • Response: User ← Proxy ← Server

– Security
  • The connection between user and proxy is encrypted.
  • Requests are shuffled before they are forwarded ⇒ no incoming request can be easily matched with any outgoing request.
  • Instead of one proxy, a cascade of proxies can be established ⇒ A single honest proxy on the route is sufficient for that no information on the user is leaked.
IP-anonymity: mix networks (2)

– Available systems
  • JAP (anon.inf.tu-dresden.de)
  • TOR (tor.eff.org)

– Major differences
  • JAP is centralized and needs funding for the proxy operators (it is going profitable soon, or it's going to die)
  • TOR has a P2P-architecture: People who want anonymity (corporations, governmental agencies, NGOs) run proxies
Market mechanisms

Data Holder \(\text{transmits } d\) Data Collector \(\text{transmits } d\) Data Analyst \(\text{transmits } f(d)\)
P3P

Platform for Privacy Preferences Project (P3P)
- developed by the W3C
- emerging as an industry standard

P3P is …
- a simple way to code privacy policies and privacy preferences
- a machine and human readable XML application
- a technical approach to privacy

Current state
- Recommendation issued 16 April 2002 (P3P 1.0)
- P3P 1.1 Specification (W3C Working Draft 1 July 2005)
P3P – main ideas

– fixing a privacy policy in a way that …
  • it can be found easily
  • it can be processed by machines
  • it can include meta-information

– users can create a privacy policy
  • express their personal privacy preferences

– web site operators can create a privacy policy
  • how does the enterprise use the provided data in the worst case
P3P – mechanisms

– The site publishes a policy reference file.

– The user agent (e.g. the browser) fetches this reference file.

– The user agent fetches the right privacy policy file.
Policy Reference File (example)

```xml
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY-REFERENCES>
    <EXPIRY max-age="172800"/>
    <POLICY-REF about="/P3P/Policies.xml#first">
      <INCLUDE>*</INCLUDE>
      <EXCLUDE>/catalog/*</EXCLUDE>
      <EXCLUDE>/cgi-bin/*</EXCLUDE>
      <EXCLUDE>/servlet/*</EXCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#second">
      <INCLUDE>/catalog/*</INCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#third">
      <INCLUDE>/cgi-bin/*</INCLUDE>
      <INCLUDE>/servlet/*</INCLUDE>
      <EXCLUDE>/servlet/unknown</EXCLUDE>
    </POLICY-REF>
  </POLICY-REFERENCES>
</META>
```
Additional requirements

- Non-ambiguity

- Multiple languages

- “Safe zone” for fetching policies
P3P – Privacy Policy

– Main content
  • meta-information
  • information about the issuer
  • a set of usage scenarios

– Each scenario (expressed by a statement) indicates
  • who (recipient)
  • performs which action (purpose)
  • on which data (data-group)
  • and how long the data will be kept (retention)

– A scenario is a potential action.
Privacy Policy File (fragment)

<STATEMENT>
  <CONSEQUENCE>
    We use this information when you make a purchase.
  </CONSEQUENCE>
  <PURPOSE><current/></PURPOSE>
  <RECIPIENT><ours/></RECIPIENT>
  <RETENTION><stated-purpose/></RETENTION>
  <DATA-GROUP>
    <DATA ref="#user.name"/>
    <DATA ref="#user.home-info.postal"/>
    <DATA ref="#user.home-info.telecom.telephone"/>
    <DATA ref="#user.business-info.postal"/>
    <DATA ref="#user.business-info.telecom.telephone"/>
    <DATA ref="#user.home-info.online.email"/>
    <DATA ref="#user.login.id"/>
    <DATA ref="#user.login.password"/>
    <DATA ref="#dynamic.miscdata">
      <CATEGORIES><purchase/></CATEGORIES>
    </DATA>
  </DATA-GROUP>
</STATEMENT>
P3P – discussion

- Main advantages
  - open standard
  - emerging standard
  - extensible standard

- Some disadvantages
  - Still limited in functionality
  - Legal restrictions and inference problems not yet accounted for

- Is P3P a PET?
APPEL

- APPEL
  - A P3P Preference Exchange Language 1.0
  - Developed in 2002, currently no active development
  - http://www.w3.org/TR/P3P-preferences

- Scope
  - APPEL is a language “for describing collections of preferences regarding P3P policies between P3P agents”.
  - A user’s preferences can be expressed by set of rules, called ruleset. Given these rulesets, parties can take decisions whether or not to accept an issued P3P privacy policy.
P3P – future development

– Notice and Communication
  • e.g. better visualization of privacy policies and possible privacy infringements

– Choice and Control
  • User agents should
    - offer configuration tools for preferences customization
    - allow users to import and customize P3P preferences from trusted parties
    - present configuration options to users in a way that is neutral or biased towards privacy
  • Service Providers should offer different privacy levels
Market mechanisms

Data Holder \(\rightarrow\) Data Collector \(\rightarrow\) Data Analyst

\[ f(d) \]

\[ \text{transmits } d \]

\[ \text{transmits } d \]

P3P
Market mechanisms

Data Holder 1

transmits \( d \)

Data Collector

transmits \( d \)

transmits \( f(d) \)

Data Analyst

Data Holder 2

Data Holder 3
Necessity of negotiations

- “Negotiation is an interpersonal decision-making process necessary whenever we cannot achieve our objectives single-handedly.”
  [Thompson, 2005]

- Negotiations can conciliate conflicting preferences:
  - Users want to keep their privacy
  - Service Providers want to process personal data for their personalization efforts
Data collection and storage is not bad per se (1)

– size and diversity of the web makes it increasingly difficult to find what one is looking for
  • not possible to examine all available alternatives
  • searching a large catalogue can be frustrating and unproductive
  • problem is exacerbated when the user
    - cannot articulate specific properties of the item (s)he is seeking
    - the user does not know exactly what (s)he is seeking

[Palmer 2002], [Parsons, Ralph, Gallagher 2004]
Data collection and storage is not bad per se (2)

– automated mechanisms have tremendous potential to help users locate desired information and/or products
  • online users have limited patience for locating material in a large information space that does not provide effective guidance
  • applying (automated) tools that
    - improve the product/consumer match (decision quality) without increasing search time (cognitive effort)
    - decrease search time without worsening the product/consumer match, may improve customer satisfaction

– user-adaptive systems rely on large data bases

[Palmer 2002], [Parsons, Ralph, Gallagher 2004]
Data collection and storage is not bad per se (3)

– empirical evidence:
  • 51% would share personal information in exchange for better service (N = 4500)
  • 73% of consumers find it helpful and convenient when a web site “remembers” basic information about them
  • 62% of users say they dislike web sites requesting personal information that they had already provided.

– but:
  • 58% of users require a privacy statement on a web site before sharing personal information

[Personalization Consortium 2000]
Privacy – Personalization trade-off

- **Personalization**
  - Better service
  - Better matching
  - Individualization
  - More convenience
  - More safety

- **Privacy loss**
  - Long-term privacy loss
  - Manipulation
  - Price discrimination
  - Identity theft
  - Sharing data

- **Benefits are clear and direct**
- **Costs are vague and indirect**
  - Expected future costs
Using P3P for Privacy Negotiations

- P3P based solutions suffer from two major drawbacks:
  - “take-it-or-leave-it”-principle
  - “one-size-fits-all”-principle

- P3P has an explicitly built-in extension mechanism.

- APPEL is not suitable for privacy negotiations:
  - no support for negotiations in the language semantics
  - no extension mechanism
Negotiable privacy dimensions

- Negotiating about entire privacy policies is not feasible.
  - Identifying relevant and negotiable privacy dimensions.

- Privacy dimension:
  - One facet of the multi-dimensional concept “User privacy”.
  - For each dimension, ordered revelation levels exist.

- Top-level privacy dimensions derived from P3P:
  - RECIPIENT: disclosure to third-parties
  - RETENTION: how long the data will be kept
  - PURPOSE: extent of use
  - DATA: detail level of the collected data

[Preibusch, 2005, 2006]
Modelling the negotiation: [Preibusch, 2005, 2006]

Model Parameters

- 2 parties:
  - Service Provider maximizes its profit
  - User maximizes its utility (U)

- Parameters:
  - $D^n$: n-dimensional privacy space
  - $d_i \in D$: privacy dimension
  - $a_i$: user’s revelation level on $d_i$
  - $a_i^T$: revelation threshold
  - $\alpha_i$: individual weighting of dimension $\alpha_i$
  - $\gamma$: user’s global privacy sensitivity
  - $R(a)$: discount provided by the service provider
  - $P(a)$: non-monetary personalization benefits
  - $B$: base utility by the execution of the contract
Modelling the negotiation: Model Properties

– **User’s utility function:**
  
  \[ U() = -\gamma \cdot \prod_{i=1}^{n} a_i^{a_i} + P(a_1, ..., a_n) + R(a_1, ..., a_n) + B \]

– **Participation constraints:**
  
  \[ R(a_1, ..., a_n) = 0 \quad \land \quad B = 0 \iff \exists i: a_i < a_i^T \]

  \[ P(a_1, ..., a_n) < -\gamma \cdot \prod_{i=1}^{n} a_i^{a_i} \quad \Rightarrow \quad \text{unpersonalized usage preferred} \]

– **Influences of the revelation levels } a_i:}
  
  \[ \frac{\partial U}{\partial a_i} \leq 0: \text{higher revelation reduces the utility} \]
  
  \[ \frac{\partial U}{\partial R} > 0: \text{the user appreciates discounts} \]
  
  \[ \frac{\partial R}{\partial a_i} \geq 0: \text{higher revelation leads to higher discounts} \]
  
  \[ \frac{\partial P}{\partial a_i} \geq 0: \text{higher revelation improves the personalization quality} \]
  
  \[ \frac{\partial B}{\partial a_i} = 0: \text{no impact of the revelation level on the base utility} \]
### Positive social impact

<table>
<thead>
<tr>
<th>Seller</th>
<th>Buyer</th>
<th>Benefits $\leq 0$</th>
<th>Benefits $&gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No transaction</td>
<td>No transaction</td>
</tr>
<tr>
<td>Profit $\leq 0$</td>
<td>Total utility $\leq 0$</td>
<td>Total utility $= ?$</td>
<td></td>
</tr>
<tr>
<td>Profit $&gt; 0$</td>
<td>Total utility $= ?$</td>
<td>Total utility $&gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>

*Transaction*
Example: Multichannel Retailing

- A multichannel retailer with a store locator.

- Privacy dimensions:
  - 2 privacy dimensions: \( d_1 \): age, \( d_2 \): address
  - \( d_1 \): revelation levels:
    - \{year (Y), year, month (YM), year, month, day (YMD)\}
  - \( d_2 \): revelation levels:
    - \{city (C), city, ZIP (CZ), city, ZIP, street (CZS)\}
  - \( a_1^T \): revelation threshold: year (Y)
  - \( a_2^T \): revelation threshold: city and ZIP (CZ) [Preibusch, 2006]
Example: Multichannel Retailing

– Possible contracts:

\[\text{d}_1 / \text{age} \]

\[\text{d}_2 / \text{address} \]

- CZS
- CZ
- C

Y YM YMD
Example: Multichannel Retailing

- Introducing data revelation thresholds:

\[ d_1 / \text{age} \]

\[ d_2 / \text{address} \]

\[ a_1^T: \text{year} \]

\[ a_2^T: \text{city + ZIP} \]

Y, YM, YMD
Example: Multichannel Retailing

- The contract space is restricted:

![Diagram with axes and annotations]
Example: Multichannel Retailing

Possible negotiation outcomes:

- d₁ / age
- d₂ / address

CZS
CZ
C
Y
YM
YMD
Example: Multichannel Retailing

– Introducing user’s iso-utility curves:
Example: Multichannel Retailing

– The Service Provider’s strategy: $D^n \rightarrow \text{ran}(R)$:
Coding negotiable contracts in P3P

– The (data, discount)-tuples can be coded in P3P.
  • offer = tuple (asked data, granted discount)

– Two new elements proposed as an EXTENSION:
  • NEGOTIATION-GROUP-DEF
    abstract pool of alternative, negotiable contracts
  • NEGOTIATION-GROUP
    nested within a STATEMENT, indicates membership to negotiation-group-def

– In accordance with syntax and semantics of STATEMENT-GROUP-DEF and STATEMENT-GROUP, as of P3P 1.1
Example of coded negotiation (fragment)

```
<EXTENSION optional="no">
  <NEGOTIATION-GROUP-DEF id="delivery" short-description="Choosing medium"/>
</EXTENSION>

...

<STATEMENT>
<EXTENSION optional="no">
  <NEGOTIATION-GROUP id="delivery"
     name="delivery as e-book" benefits="10% discount"/>
</EXTENSION>
<DATA-GROUP>
  <DATA ref="#user.home-info.online.email"/>
</DATA-GROUP>
</STATEMENT>

...

<STATEMENT>
<EXTENSION optional="no">
  <NEGOTIATION-GROUP id="delivery"
     name="delivery as hard copy" benefits="robust hard-cover"/>
</EXTENSION>
<DATA-GROUP>
  <DATA ref="#user.name"/>
  <DATA ref="#user.home-info.postal"/>
</DATA-GROUP>
</STATEMENT>
```
Example of coded negotiation (fragment)

```xml
<EXTENSION optional="no">
  <NEGOTIATION-GROUP-DEF id="delivery" short-description="Choosing medium" />
</EXTENSION>

...

<STATEMENT>
  <EXTENSION optional="no">
    <NEGOTIATION-GROUP id="delivery"
      name="delivery as e-book" benefits="10% discount" />
  </EXTENSION>
  <DATA-GROUP> <DATA ref="#user.home-info.online.email"/> </DATA-GROUP>
</STATEMENT>

...

<STATEMENT>
  <EXTENSION optional="no">
    <NEGOTIATION-GROUP id="delivery"
      name="delivery as hard copy" benefits="robust hard-cover" />
  </EXTENSION>
  <DATA-GROUP> <DATA ref="#user.name"/> <DATA ref="#user.home-info.postal"/> </DATA-GROUP>
</STATEMENT>
```
Example of coded negotiation (fragment)

```xml
<EXTENSION optional="no">
  <NEGOTIATION-GROUP-DEF id="delivery" short-description="Choosing medium"/>
</EXTENSION>

...

<STATEMENT>
<EXTENSION optional="no">
  <NEGOTIATION-GROUP id="delivery"
    name="delivery as e-book" benefits="10% discount"/>
</EXTENSION>
<DATA-GROUP>
  <DATA ref="#user.home-info.online.email"/>
</DATA-GROUP>
</STATEMENT>

...

<STATEMENT>
<EXTENSION optional="no">
  <NEGOTIATION-GROUP id="delivery"
    name="delivery as hard copy" benefits="robust hard-cover"/>
</EXTENSION>
<DATA-GROUP>
  <DATA ref="#user.name"/>
  <DATA ref="#user.home-info.postal"/>
</DATA-GROUP>
</STATEMENT>
```
Market mechanisms

Data Holder 1 transmits \( d_1 \),

Data Collector transmits \( d_1, d_2, d_3 \),

Data Analyst transmits \( f(d) \),

Data Holder 2 transmits \( d_2 \),

Data Holder 3 transmits \( d_3 \).
Multiple privacy policies

- Negotiations lead to individual privacy policies
- Every data has specific usage restriction
- Sticky Policy Paradigm
  \[(P, \{d_1, d_2, d_3\}) \rightarrow \{(P_1, d_1), (P_2, d_2), (P_3, d_3)\}\]
- Adequate aggregation semantics needed
Enterprise Privacy Policies

– EPAL policy
  • EPAL defines policy terminology and authorization rules
  • Rules allow/deny privacy relevant actions, depending on purpose
  • Can be mapped to P3P, supporting consistent internal/external views

– External privacy promises must be enforced
EPAL – example

- Privacy promise:
  - "Email can only be used for the book-of-the-month club if consent has been given and age is more than 13":

- EPAL rule:
  
  ```xml
  <ALLOW
    data-user="borderless-books"
    data-category="email"
    purpose="book-of-the-month-club"
    operation="read"
    condition="/CustomerRecord/Consent/BookClub=True
    && /CustomerRecord/age>13"/>
  ```
EPAL semantics

– Inheritance:
  • “Allow” inherits down along hierarchies
  • “Deny” inherits up and down

– Deployment descriptions for abstract EPAL vocabularies

– Requirements:
  • Application-independent Policies 'CPO driven'
  • Enforce this policy across multiple applications
Market mechanisms

Data Holder 1 transmits $d_1$

Data Collector

Data Holder 2 transmits $d_2$

Data Holder 3 transmits $d_3$

Data Collector transmits $d_1, d_2, d_3$

Data Analyst transmits $f(d)$

Data Holder 1

Data Holder 2

Data Holder 3

Data Collector

Data Analyst
Market mechanisms

Data Holder 1

Data Holder 2

Data Holder 3

Data Collector

Data Analyst

transmits \( d_1 \)

transmits \( d_1, d_2, d_3 \)

transmits \( f(d) \)

\( P_{3P}^1 \)

\( P_{3P}^2 \)

\( P_{3P}^3 \)

\( P_{3P}^{agg} \)

\( d_2 \)

\( d_3 \)

\( EPAL 1 \)

\( EPAL 2 \)

\( EPAL t \)
SIMT framework [Teltzrow, Preibusch, Berendt, 2004]
Privacy policies and seals

– Privacy Policies Drawbacks
  • Difficult to understand
  • Currently only accessible to expert users
  • Hard to communicate

– Privacy seals
  • Bringing trust to privacy policies
  • Drawbacks: declarative level, only minor compliance checks
Vision: privacy verification

– Automated compliance checks

  • for data collection practices
    - XForms allow forms to be annotated by P3P data types

  • for data processing practices
    - Process algebra and refinement analyses
Future menaces of Privacy: RFID

– RFID used to …
  • collect information linked to personal data
  • store personal data on each tag
  • track without traditional identifiers being available

– Privacy threats
  • Leaking information pertaining to personal property
  • Tracking the user’s transaction history

– Fair information practices undermined by RFID
Possible answers to RFID

– Kill the tag functionality
  • Curtails the future potential use of RFID in consumer services
  • Prevents the use of RFID for reselling or recycling

– Trade-off
  • Privacy level
  • Scalability
  • Costs of tag

– No definitive answer yet
Thank you!

– Remaining questions and Discussion

– Sören Preibusch
  German Institute for Economic Research (DIW)
  Königin-Luise-Str. 5
  14191 Berlin, Germany
  spreibusch@diw.de

– Extended version available online soon