TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Computer Science Institute of Systems Architecture, Privacy and Data Security Research Group

Valuating Privacy with Option Pricing Theory

WEIS 2009

Stefan Berthold and Rainer Böhme

{stefan.berthold, rainer.boehme} @ tu-dresden.de

University College London, 25 June 2009

Structure

1. Problem statement

2. Privacy options

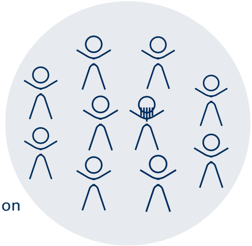
3. Modelling uncertainty
   - Timed linkability process (micro model)
   - Population development (macro model)

4. Valuation methods

TECHNISCHE UNIVERSITÄT DRESDEN

Motivating Example

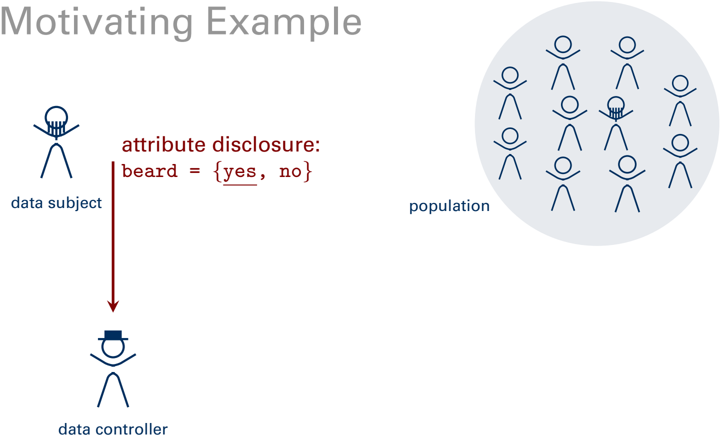
data subject



population

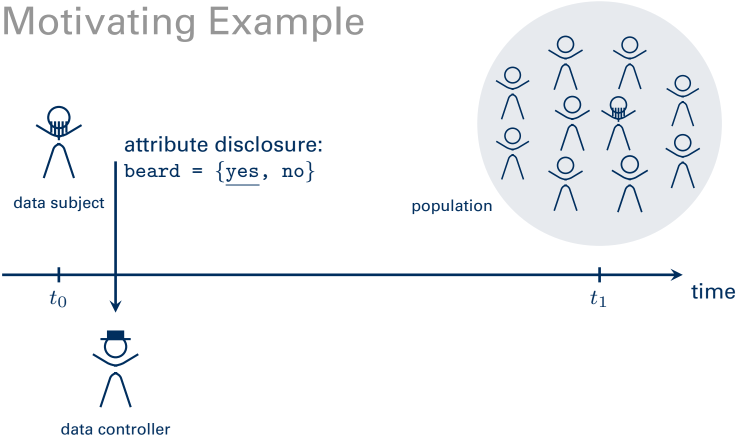
TECHNISCHE UNIVERSITÄT DRESDEN

Motivating Example



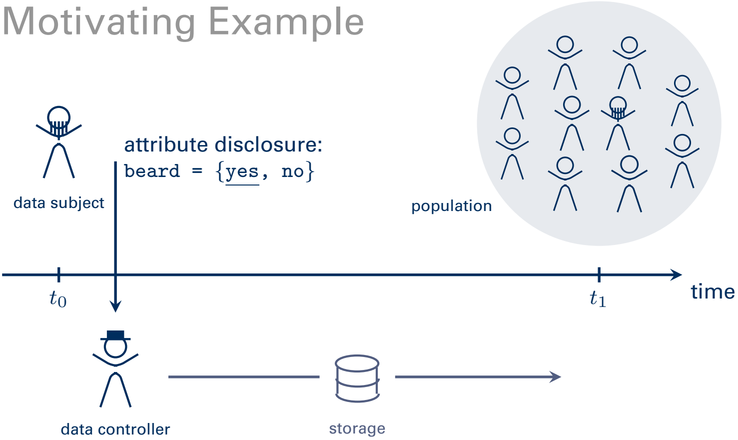
Simplification: single, binary, and perfectly observable attribute

Motivating Example



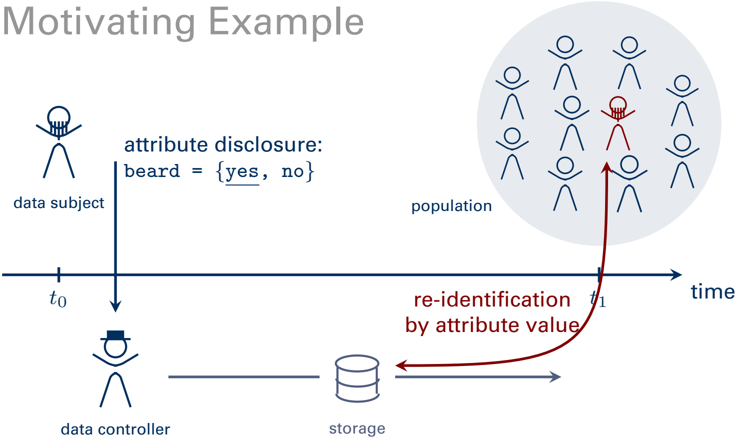
Simplification: single, binary, and perfectly observable attribute

Motivating Example



Simplification: single, binary, and perfectly observable attribute

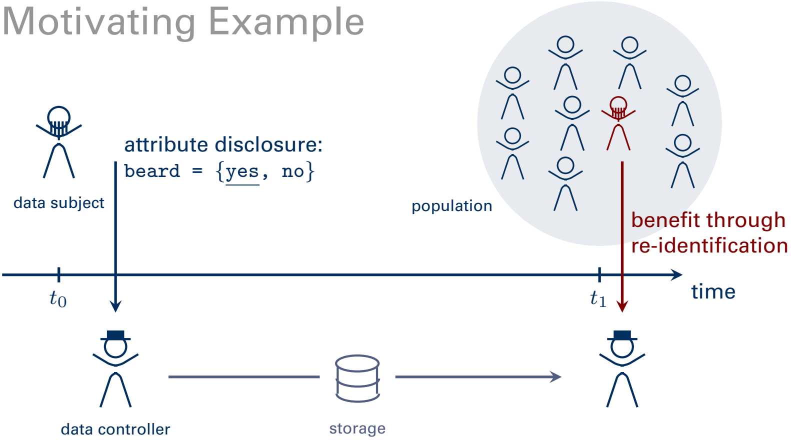
Motivating Example



Simplification: single, binary, and perfectly observable attribute

TECHNISCHE UNIVERSITÄT DRESDEN

Motivating Example



Simplification: single, binary, and perfectly observable attribute

Research Question

How can we measure the long-term value / cost of personal data disclosure?

Research Question

privacy metrics

How can we measure the long-term value / cost of personal data disclosure?

Research Question

untackled

How can we measure the long-term value / cost of personal data disclosure?

Research Question

How can we measure the long-term

value / cost of personal data disclosure?

application: decision-support in future
privacy-enhancing technologies (PET)

TECHNISCHE UNIVERSITÄT DRESDEN

Research Question

How can we measure the long-term

value / cost of personal data disclosure?

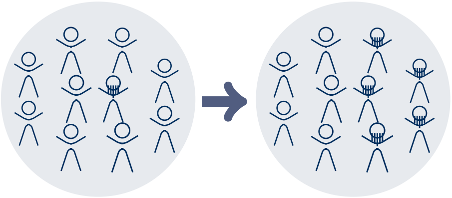
business perspective:
valuation of customer databases

Sources of Uncertainty

Micro-level

behaviour of data subject

Macro-level



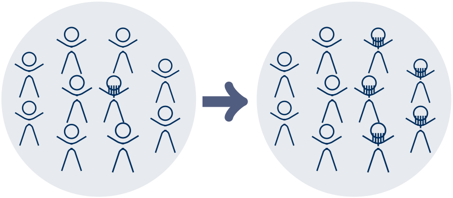
high self-information low self-information

uniqueness of attribute value in the population

Sources of Uncertainty

Micro-level

Macro-level



high self-information

low self-information

behaviour of data subject

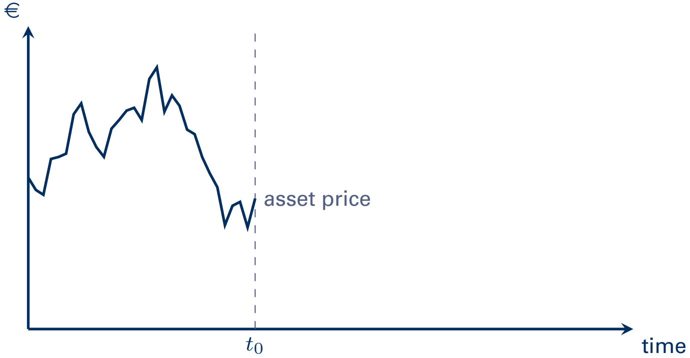
uniqueness of attribute value in the population

Approach: Borrow from financial maths to model uncertainty over time.

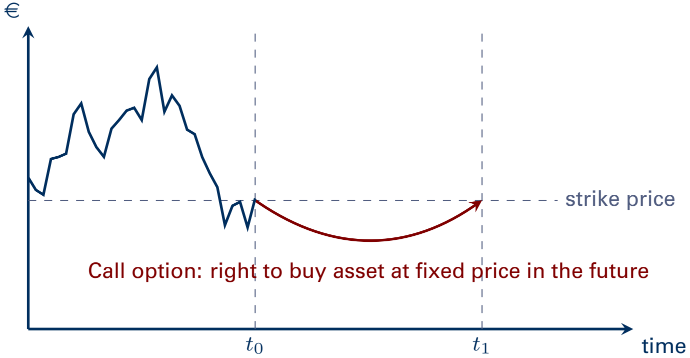
2

Privacy Options

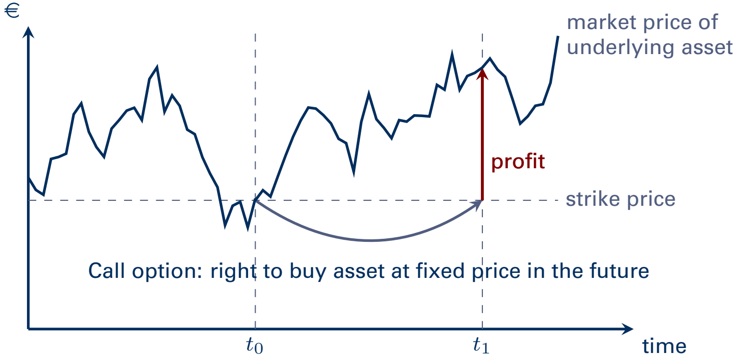
Recall: Financial Options



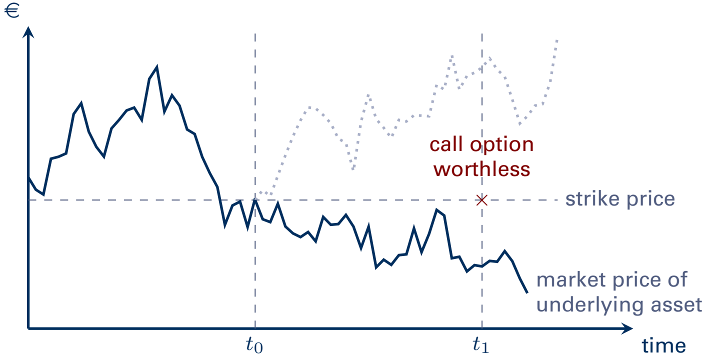
Recall: Financial Options



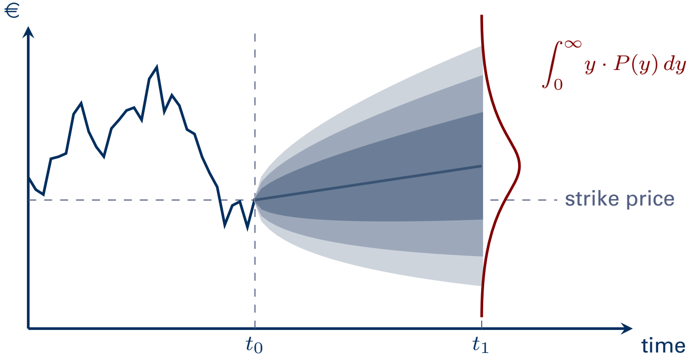
Recall: Financial Options



Recall: Financial Options



Recall: Financial Options



Option value: Probability-weighted average of all possible outcomes at t_1.

Black, Scholes, Merton (1970s)

Analogy

money = information

in Shannon's sense

Analogy

money = information

in Shannon's sense

€ = bits

unit of information needed
for re-identification

Privacy Options

Personal data disclosure is like writing a call option: the counterpart obtains information that can be used for re-identification in the future.

Privacy Options

strike price: cost of data retrieval

Personal data disclosure is like writing a call option: the counterpart obtains information that can be used for re-identification in the future.

TECHNISCHE UNIVERSITÄT DRESDEN

Privacy Options

Personal data disclosure is like writing a call option: the counterpart obtains information that can be used for re-identification in the future.

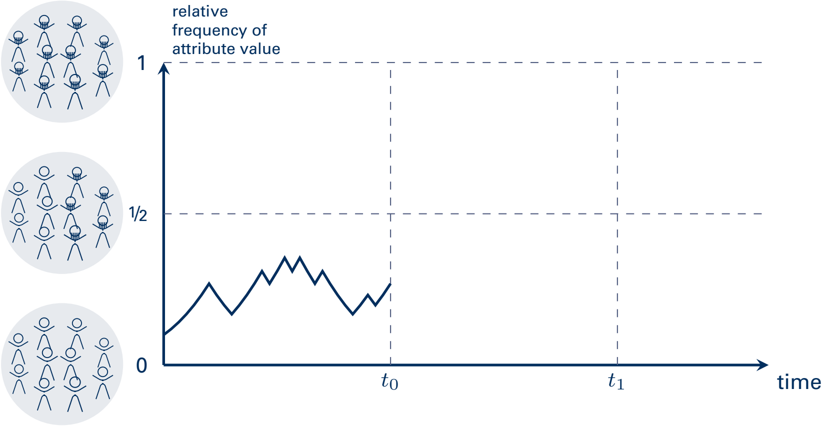
premium: incentive (e. g., rebate via loyalty card)

Privacy Options

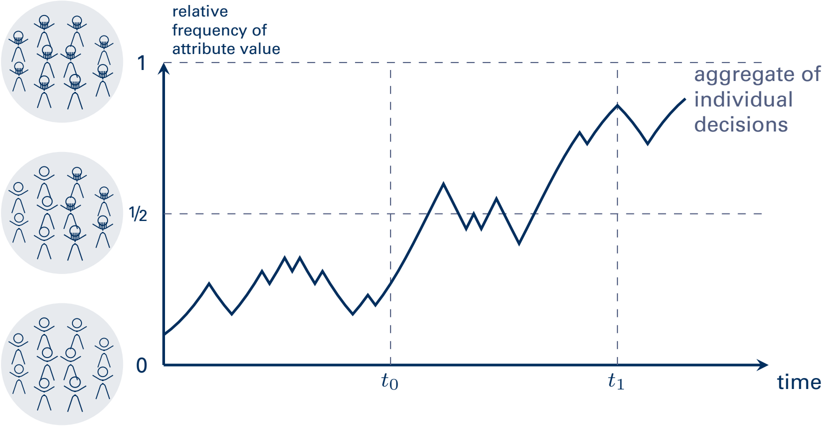
Personal data disclosure is like writing a call option: the counterpart obtains information that can be used for re-identification in the future.

value: depends on population development

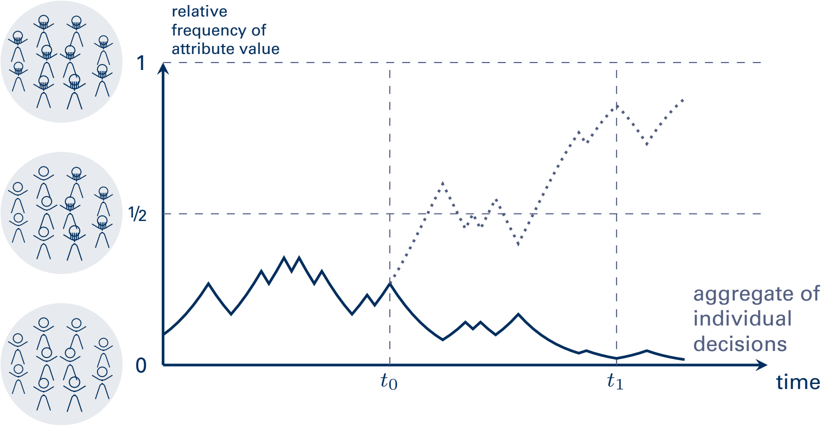
Underlying: Population Development



Underlying: Population Development



Underlying: Population Development



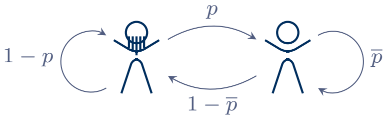
TECHNISCHE UNIVERSITÄT DRESDEN

3

Modelling Uncertainty

Timed Linkability Process

State Space Model



Initial state

$$\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

State transition matrix

$$\boldsymbol{A} = \begin{pmatrix} p & 1-\overline{p} \\ 1-p & \overline{p} \end{pmatrix}$$

(can be specific for each data subject)

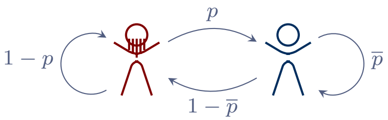
State change

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\,\boldsymbol{x}_t = \boldsymbol{A}^{t+1}\,\boldsymbol{x}_0$$

TECHNISCHE UNIVERSITÄT DRESDEN

Timed Linkability Process

State Space Model



Initial state

$$\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

State transition matrix

$$\boldsymbol{A} = \begin{pmatrix} p & 1-\overline{p} \\ 1-p & \overline{p} \end{pmatrix}$$

(can be specific for each data subject)

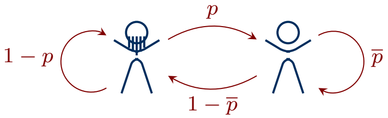
State change

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\,\boldsymbol{x}_t = \boldsymbol{A}^{t+1}\,\boldsymbol{x}_0$$

TECHNISCHE UNIVERSITÄT DRESDEN

Timed Linkability Process

State Space Model



Initial state

$$\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

State transition matrix

$$\boldsymbol{A} = \begin{pmatrix} p & 1-\overline{p} \\ 1-p & \overline{p} \end{pmatrix}$$

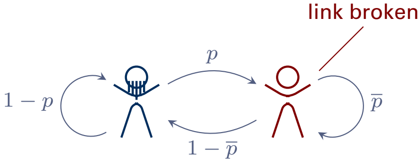
(can be specific for each data subject)

State change

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{A}\,\boldsymbol{x}_t \\ &= \boldsymbol{A}^{t+1}\,\boldsymbol{x}_0 \end{aligned}$$

Timed Linkability Process

State Space Model



Initial state

$$\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

State transition matrix

$$\boldsymbol{A} = \begin{pmatrix} p & 1-\overline{p} \\ 1-p & \overline{p} \end{pmatrix}$$

(can be specific for each data subject)

State change

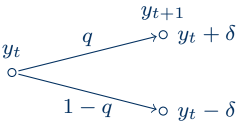
$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{A}\,\boldsymbol{x}_t \\ &= \boldsymbol{A}^{t+1}\,\boldsymbol{x}_0 \end{aligned}$$

Population Development

Binomial Random Walk

Similar to underlying process in Binomial Option Pricing Model (BOPM)

Cox, Ross and Rubinstein, 1979



q : trend

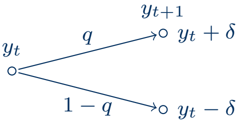
δ : dispersion (volatility)

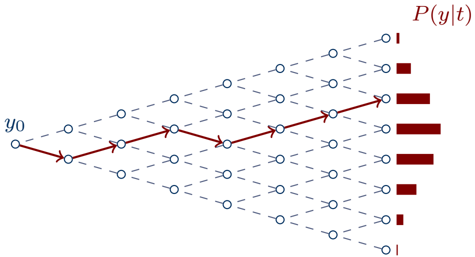
Population Development

Binomial Random Walk

Similar to underlying process in Binomial Option Pricing Model (BOPM)

Cox, Ross and Rubinstein, 1979





q : trend

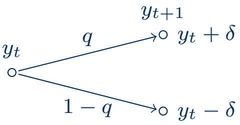
δ : dispersion (volatility)

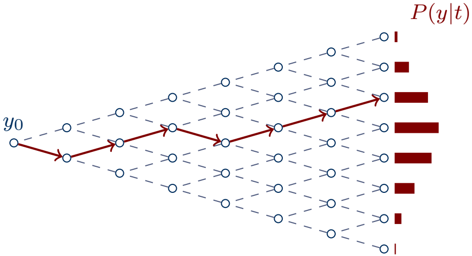
Population Development

Binomial Random Walk

Similar to underlying process in Binomial Option Pricing Model (BOPM)

Cox, Ross and Rubinstein, 1979





q : trend

δ : dispersion (volatility)

Inverse logit transform: random walk domain $\mathcal{Y} \rightarrow$ probability dom. $\mathcal{R} \subset [0, 1]$

4

Valuation Methods

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} \quad = \quad P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

TECHNISCHE UNIVERSITÄT DRESDEN

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} \quad = \quad P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

probability of attribute value v observed at t_0 remaining valid until t_1.

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} \quad = \quad P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

probability of attribute value v observed at t_0 remaining valid until t_1.

$\longrightarrow$ state space model

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} = P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

probability of attribute value v observed at t_0 remaining valid until t_1.

→ state space model

expected self-information of attribute value v at t_1.

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} = P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

probability of attribute value v observed at t_0 remaining valid until t_1.

→ state space model

expected self-information of attribute value v at t_1.

$$\mathcal{H}_v(t) = -\sum_{r \in \mathcal{R}_v} P(r|t) \log_2 r$$

Shannon 1948, cf. Equation (12)

Baseline Valuation Method

Combination of both stochastic processes and basic information theory:

$$\text{value of data disclosure} = P(v|t_1) \cdot \mathcal{H}_v(t_1)$$

probability of attribute value v observed at t_0 remaining valid until t_1.

→ state space model

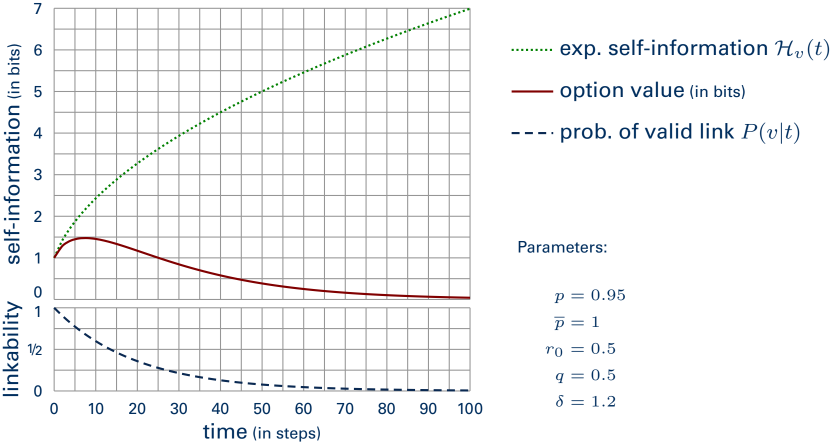
expected self-information of attribute value v at t_1.

$$\mathcal{H}_v(t) = -\sum_{r \in \mathcal{R}_v} P(r|t) \log_2 r$$

→ population development

Shannon 1948, cf. Equation (12)

Numerical Example



Parameters:

$$p = 0.95$$
$$\overline{p} = 1$$
$$r_0 = 0.5$$
$$q = 0.5$$
$$\delta = 1.2$$

Variants

	Aggregation over time	
Future value of ...	point in time t_1	time range $t_0 \ldots t_1$
... a specific data subject's attribute value (self-information)	$P(v \vert t_1) \cdot \mathcal{H}_v(t_1)$ cf. Equation (13)	$\propto \sum_{t=0}^{t_1} P(v \vert t) \cdot \mathcal{H}_v(t)$ cf. Equation (15)

Variants

	Aggregation over time	
Future value of . . .	point in time t_1	time range $t_0 \ldots t_1$
. . . a specific data subject's attribute value (self-information)	$P(v\|\,t_1) \cdot \mathcal{H}_v(t_1)$ cf. Equation (13)	$\propto \sum_{t=0}^{t_1} P(v\|\,t) \cdot \mathcal{H}_v(t)$ cf. Equation (15)
. . . an attribute of a random individual (entropy)	$-\sum_{v \in \mathcal{V}} P(v\|t_0 \wedge t_1) \cdot \mathcal{H}_v(t_1)$	$\propto -\sum_{t=0}^{t_1} \sum_{v \in \mathcal{V}} \cdots$

Variants

Future value of ...	Aggregation over time: point in time t_1	Aggregation over time: time range $t_0 \ldots t_1$
... a specific data subject's attribute value (self-information)	$P(v\| t_1) \cdot \mathcal{H}_v(t_1)$ cf. Equation (13)	$\propto \sum_{t=0}^{t_1} P(v\| t) \cdot \mathcal{H}_v(t)$ cf. Equation (15)
... an attribute of a random individual (entropy)	$-\sum_{v \in \mathcal{V}} P(v\|t_0 \wedge t_1) \cdot \mathcal{H}_v(t_1)$	$\propto -\sum_{t=0}^{t_1} \sum_{v \in \mathcal{V}} \cdots$
	↓ "European" option	↓ "American" option

Summary: Options Zoo

financial option Right to buy (resp. sell) security on financial market at defined price in the future.

→ eliminate market risk via hedging

Summary: Options Zoo

financial option Right to buy (resp. sell) security on financial market at defined price in the future.

→ eliminate market risk via hedging

real option Transfer the idea of financial options to tangible investments, e. g., project management: possibility to decide about an investment project in the future.

→ account for mid-course strategy corrections

Summary: Options Zoo

	financial option	Right to buy (resp. sell) security on financial market at defined price in the future. → eliminate market risk via hedging
	real option	Transfer the idea of financial options to tangible investments, e. g., project management: possibility to decide about an investment project in the future. → account for mid-course strategy corrections
NEW:	**privacy option**	Combine the idea of financial options with information theory: knowledge of personal data allows the counterpart to (better) identify a data subject in the future. → model the value of personal data

Concluding Remarks

Contribution

Novel interpretation of data disclosure as writing a call option

Next steps

- Generalisation to multiple multi-valued attributes
  (under submission)
- Composition of options in privacy policy languages
  cf. for financial options: Peyton Jones and Eber, 2003

Limitations

- Epistemic: model mismatch and incomplete information
- Behavioural: bounded rationality and time-inconsistent discounting
  Acquisti and Grossklags, 2005

Q & A

Thank you for your attention.

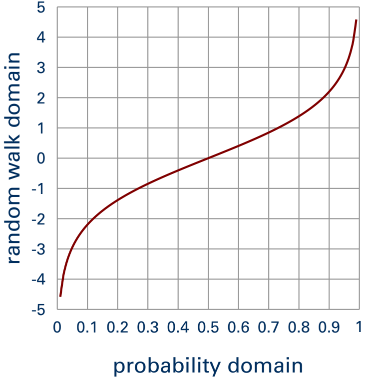
Stefan Berthold and Rainer Böhme

{stefan.berthold, rainer.boehme} @ tu-dresden.de

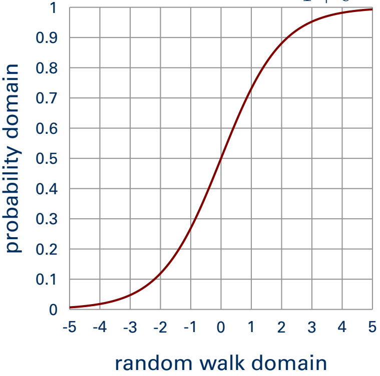
University College London, 25 June 2009

Logit and Inverse Logit Transformation

$$\text{logit}(x) = \log \frac{x}{1-x}$$



$$\text{logit}^{-1}(x) = \frac{e^x}{1+e^x}$$



Related Work

Privacy metrics

Entropy-based anonymity metrics
Díaz et al., 2002; Serjantov and Danezis, 2002; Tóth et al., 2004

Generalisation to (un)linkability
Pfitzmann and Hansen, 2008; Steinbrecher and Köpsell, 2003; Clauß, 2006

Empirical measurement of privacy
e. g.: Huberman et al., 2005; Berendt et al., 2005; Grossklags and Acquisti, 2007

Financial methods in information security

Security investment
Soo Hoo, 2002; Gordon and Loeb, 2002; Purser, 2004

Real options for information security investment
Gordon et al., 2003; Daneva, 2006; Li and Su, 2007; Herath and Herath, 2008; Tatsumi and Goto, 2009

Market-mechanism for security indicators
Matsuura, 2001; Schechter, 2004; Ozment, 2004; Böhme, 2006