Designing Systems with Privacy: formal and experimental methods

Giuseppe Primiero

Department of Computer Science
Middlesex University, London
www.cs.mdx.ac.uk/people/giuseppe-primiero/

Politecnico di Milano

A reminder: Methodologies for CS

1. The mathematical trend: the influence of logic and complexity on the notion of calculation, algorithm and program;
2. The engineering trend: the construction of physical devices to perform automated tasks;
3. The scientific trend: the use of machines to perform scientific tasks.

Research Questions

- How do key human values manifest themselves in socio-technical systems?
- How can the different methodologies help in identifying such values and their roles?
- We have considered a combined methodology for Trust, focus today on Privacy

Todays' Tasks and Conceptual Ingredients

- We offer a model of IS design that accounts for a formal definition of privacy
- We work on this definition through an engineering/experimental setting
- The type of issues related to the notions of model, experiment and validation are different than the case of simulation analysed on Monday.

Privacy

- Privacy has become under threat with the rise of ITs.
- It entails both freedom of intrusion and control of personal information.
  - software systems design
  - databases
  - often non-functional (rather motivated by commercial use)
  - social networks (design to trade away personal information)
  - IoT, smartphones, etc.

Designing computational systems with Privacy, [Warnier et al., 2015]

1. Policy of no personal data storage (impractical)
2. Strict rules and best practices for storing
   - transparency: what
   - purpose: what for
   - proportionality: how much
   - access: instructions for correcting errors
   - transfer: explicit permission
   - Privacy by Design / VSD
3. Only anonymized data can be stored
4. Monitoring implemented to help best practices from the users

Privacy Online: Pre-History

- 1978, Bulletin Board System (Ward Christensen, Randy Suess) used encryption schemes for data protection
- 1980, USENET (Tom Truscott, Jim Ellis), SSL (Secure Socket layer) encryption for user identification protection
- 1994, Geocities received complaint from FTC (Federal Trading Commission) for violating privacy promises on users data collection

Privacy Online: Social Networks

- voluntary information disclosure [Gross and Acquisti, 2005] [Irani et al., 2011],
- incomplete personal information [Xu et al., 2008]
- control on web-tracking activities [Takano et al., 2014]
- socio-demographic considerations [Houghton and Joinson, 2010] [Hazari and Brown, 2014] [Jeonga and Coylea, 2014]
- design of access control [Kang and Kagal, 2010],
- requirements elicitation [Omoronyia et al., 2013]
- privacy awareness and risk assessment as privacy concerns [Tan et al., 2012]

Bayesian approaches to Privacy

- [Troncoso, 2013]: information leakage over networks
- [Gürses et al., 2008]: privacy breaches
- [Balebako and Cranor, 2014]: difficulties in defining and implementing privacy decisions
- [Kelley et al., 2013]: informed decisions on app-selection from mobile users.
- [Li et al., 2015]: conflicts between privacy configurations and network functionalities;
- [Krishnamurthy and Wills, 2008]: characterizing minimal information sets required to share for accomplishing interactions
- [Li, 2012]: evaluate the trade-offs between privacy and risks and delivers a decision method to predict the user's intention to share information online

Probabilistic methodology

- probabilistic algorithms: randomized procedures with input from a probability distribution and algorithm working for most inputs;
- probabilistic algorithms are common in recommendations and information sharing systems
- Bayesian epistemology in algorithms design reflects normative choices and creates behavioural patterns whose consequences on, for example, users' privacy and security are still only partially explored

A problem

Problem (Privacy Change Assessment)

Given a measurable amount of network activity involving user U_1 over a fixed span of time Δt, how does the probability p that some piece of information i shared by U_1 is exposed through the network to some non-connected user U_2 change over Δt?

$$PC(\Delta t) = (p(U_1 i U_2)_{t'}) - (p(U_1 i U_2)_t) \mid NA_{\Delta t}(U_1)$$

Methodological Analysis

- How can we model this case?
- How many/Which cases do we need to consider for a reliable analysis?
- What is the value of the data we monitor? (constantly changing set of data, possibly with regularities across user types)
- Which limits can be recognized in terms of evaluation?

Strategy

1. Define an axiomatic theory of informational privacy with definitions and rules
2. Formulate a version applicable to the SN context
3. Develop of tool for running experiments
4. Extract and evaluate data

Informational Privacy

[Floridi, 2005, Floridi, 2006], construed around four basic notions:

- information accessibility: the ontological features of agents and their interaction environment
- informational gap: defined by accessibility, the larger it is, the less agents know about each other;
- ontological friction: a property of the world;
- information flow: dependent on friction, holds within the system.

3 Axiomatic Laws

Axiom (Greater Gap, Greater Privacy)

$(InfoGap(A, B) > InfoGap(C, D)) \rightarrow (InfoPrivacy(A, B) > InfoPrivacy(C, D))$

3 Axiomatic Laws

Axiom (Greater Gap, Greater Privacy)

$(InfoGap(A, B) > InfoGap(C, D)) \rightarrow (InfoPrivacy(A, B) > InfoPrivacy(C, D))$

Axiom (Greater Access, Lesser Gap)

$(InfoAccess(A, B) > InfoAccess(C, D)) \rightarrow (InfoGap(A, B) < InfoGap(C, D))$

3 Axiomatic Laws

Axiom (Greater Gap, Greater Privacy)

$(InfoGap(A, B) > InfoGap(C, D)) \rightarrow (InfoPrivacy(A, B) > InfoPrivacy(C, D))$

Axiom (Greater Access, Lesser Gap)

$(InfoAccess(A, B) > InfoAccess(C, D)) \rightarrow (InfoGap(A, B) < InfoGap(C, D))$

Axiom (Greater Friction, Lesser Access)

$(OntoFriction(InfoFlow(W[A, B])) > OntoFriction(InfoFlow(W[C, D]))) \rightarrow (InfoAccess(A, B) < InfoAccess(C, D))$

A Simple Theorem

Theorem

$$(OntoFriction(InfoFlow(W[A,B])) < OntoFriction(InfoFlow(W[C,D])) \rightarrow$$

$$InfoAccess(A,B) > InfoAccess(C,D)) \rightarrow$$

$$(InfoGap(A,B) < InfoGap(C,D) \rightarrow InfoPrivacy(A,B) < InfoPrivacy(C,D))$$

Redefining the Theory for SN

Definition (Information Access)

A measure of an agent's activity on the network.

Redefining the Theory for SN

Definition (Information Access)

A measure of an agent's activity on the network.

Definition (Network Friction)

A measure of an agent's network fluidity.

Redefining the Theory for SN

Definition (Information Access)

A measure of an agent's activity on the network.

Definition (Network Friction)

A measure of an agent's network fluidity.

Redefining the Theory for SN

Definition (Information Access)

A measure of an agent's activity on the network.

Definition (Network Friction)

A measure of an agent's network fluidity.

Definition (Information Gap)

A measure of the degree of accessibility to personal data, as a function of informational access.

Redefining the Theory for SN

Definition (Information Access)

A measure of an agent's activity on the network.

Definition (Network Friction)

A measure of an agent's network fluidity.

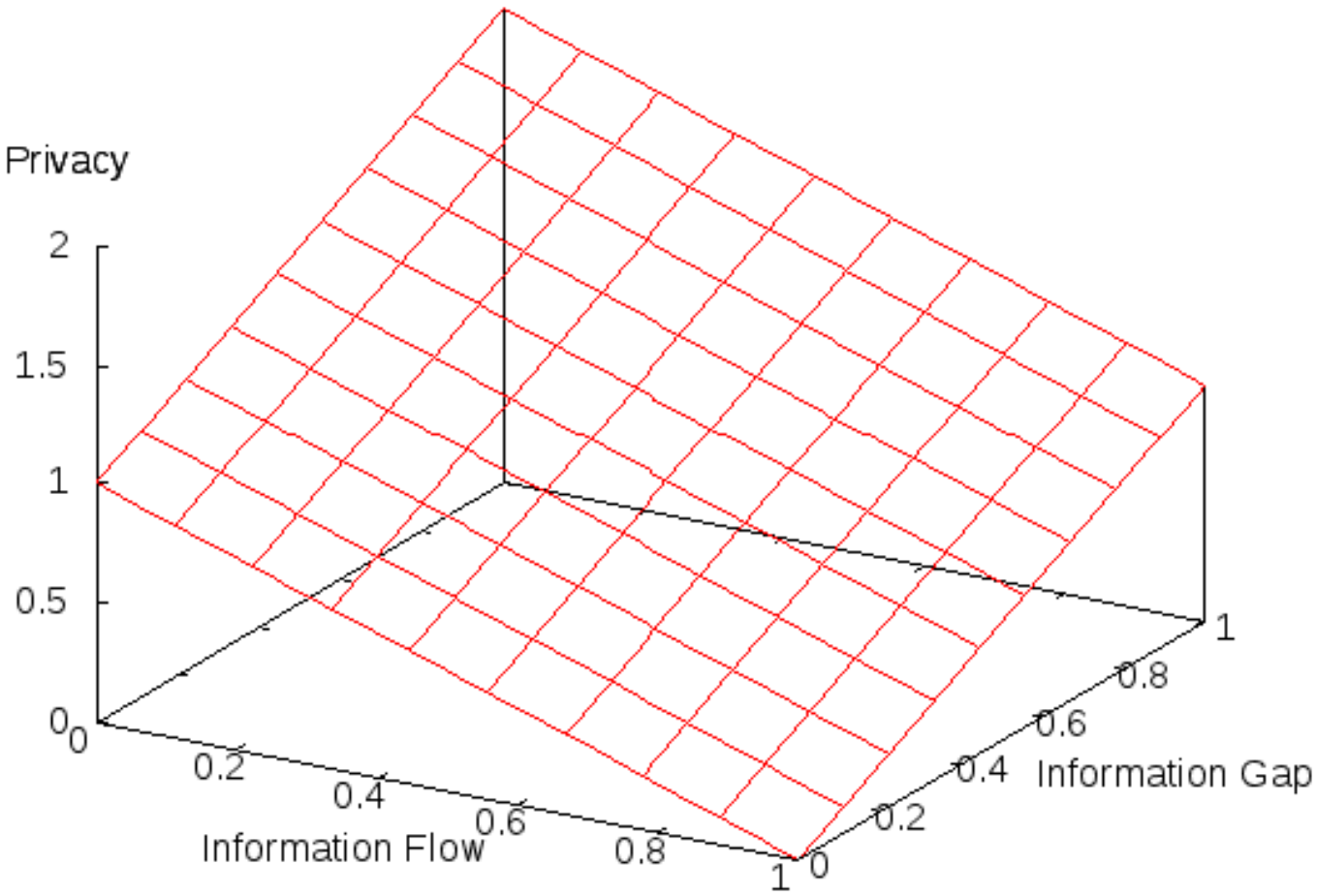
Definition (Information Gap)

A measure of the degree of accessibility to personal data, as a function of informational access.

Definition (Information Flow)

A measure of the degree of fluidity of personal data as a function of network friction.

Relations


Privacy
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Information Flow
0
0.2
0.4
0.6
0.8
1
Information Gap

Objectives for the formal analysis

- Which properties for privacy is the system able to express?
- Are the properties consistent?
- Do we gain interesting relations?
- Can various settings be modelled in this system?

The engineering implementation

- A tentative implementation of the theory
- Coded to be used on live platforms
- Experiments to be performed with real users

Java

```
// Define the first prior probability: Access
//Access is calculated in terms of
// availability of the technology
InfoAccess.setCPTable(0.30,0.70);

//Define the first dependent probability: InfoGap
//InfoGap is dependent from Access
                  //InfoAccess    high    low
InfoGap.setCPTable ("lowAccess", 0.60, 0.40);
InfoGap.setCPTable ("highAccess", 0.40, 0.60);



// Define the second prior probability: Network Friction
//Network Friction is calculated in terms of
//proportion of common nodes (over a given span of time)
//instances of actual interactions (over the same time span)
NetworkFriction.setCPTable (0.30, 0.70);


//Define the second dependent probability: Information Flow
//Information Flow is dependent from Network Friction
//Set Friction and Lubrication
                    //Friction  Frictioned    Lubricated
InfoFlow.setCPTable ("Friction", 0.70, 0.30);
InfoFlow.setCPTable ("Lubricated", 0.30, 0.70);



InfoPriv.setEquation ("InfoPriv (InfoGap, InformationFlow) = absent  || present");
InfoPriv.equationToTable (1, false, false);

net.compile();
```

Figure: Snippet of Java Implementation

Computing values

$$IP = (max(InfoGap) + min(InfoFlow))$$

$$IO = (min(InfoGap) + max(InfoFlow))$$

Java

```
gprimiero@giuseppe-laptop:~/NeticaJ_504/InfoPriv$ javac -deprecation -classpath ../bin/NeticaJ.jar BuildNet.java
gprimiero@giuseppe-laptop:~/NeticaJ_504/InfoPriv$ java -classpath .:../bin/NeticaJ.jar -Djava.library.path=../bin BuildNet

The positive value of the information gap is 0.45999998

The negative value of the information gap is 0.54

The positive value of the information flow is 0.58

The negative value of the information flow is 0.42000002

Given the max of infogap and the min of flow,
the probability of informational privacy is 0.88

Given the min of infogap and the max of flow,
the probability of informational openess is 1.12
```

Figure: Snippet of Java Implementation

Implementing definitions

Given a time interval Δt Information Access in the interval $(-t, 0)$:

$$\begin{aligned} IA(-t,0) \quad = \quad & \text{N.of likes}(-t,0) + \text{N.of locations}(-t,0) + \\ & + \text{N.of events}(-t,0) + \\ & + \text{N.of posts}(-t,0) + \text{N.of new friends(-t,0)} \end{aligned} \tag{1}$$

Implementing definitions

Network Friction in a time interval Δt as:

$$\begin{aligned} NF(-t,0) &= \text{N.of likes}(-t,0) + \text{N.of locations}(-t,0) \\ &\quad + \text{N.of events}(-t,0) + \text{N.of public posts}(-t,0) \end{aligned} \tag{2}$$

Implementing definitions

The value for Information Gap (with fixed weights) is:

$$IG = \frac{0.3NP_{\mathrm{C}} + 0.5NP_{\mathrm{AF}} + 0.8NP_{\mathrm{FofF}} + NP_{\mathrm{No}} + NP_{\mathrm{Ev}}}{\text{N.of posts}(-t, 0)} \tag{3}$$

Implementing definitions

The value for Information Flow is:

$$IF = \frac{Li + E + Sh + NP_{\mathrm{No}} + NP_{\mathrm{Ev}}}{NP_{\mathrm{Pr}} + NP_{\mathrm{C}} + NP_{\mathrm{AF}} + 0.5NP_{\mathrm{FofF}}} \tag{4}$$

Tests

Case 1. *A software developer using Facebook to interact with friends and family. The overall activity of the user has decreased due to an increased workload and to a house move. The total number of posts decreased from 31 to 16 and the number of likes decreased from 15 to 6. As expected, the value of privacy has increased to 1.32.*

Tests

Case 2. An academic using Facebook mainly to communicate research, interact with students and with some family members. Also in this case, the number of posts decreased from 124 in the period 2013-14 to 94 in 2014-15. However, the number of friends posting on this person's wall increased from 12 to 30 and, moreover, the number of shares increased from 10 to 28. As a result, the privacy of the user decreased from 1 to 0.69 (see Figure 4). This case shows that, in our model for Facebook, social network privacy is affected not only by direct user's choices, but also by the behaviour of someone's friends.

Tests

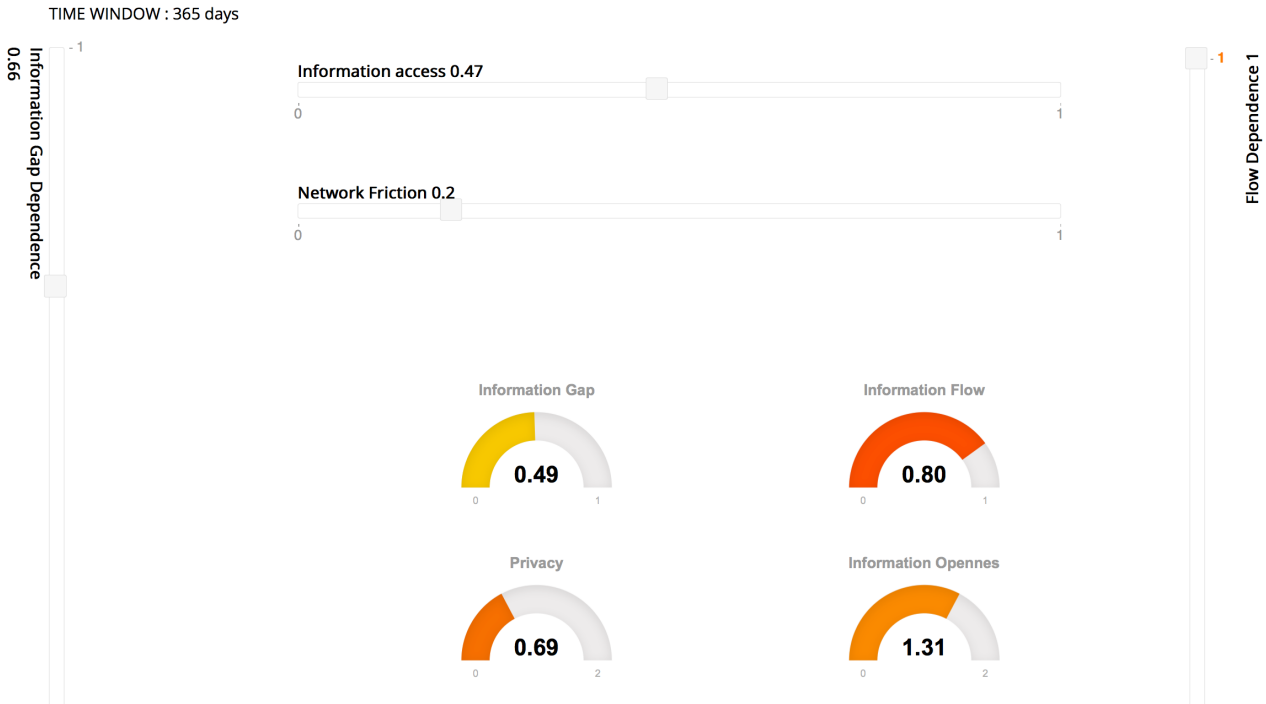


Figure: Facebook plug-in: results for User 2.

Tests

Case 3. A PhD student, using Facebook mainly to keep in touch with friends and family during her study periods abroad. For this user, in the given period, the total number of tagged places has increased from 36 to 78, and the total number of events from 50 to 57. As a result, the privacy of the user has decreased from 1 to 0.75 (see Figure 5). This case shows that, in our model, social network privacy takes into account also how much the user exposes information on her non-digital ontology, through the mention of places and events.

Tests

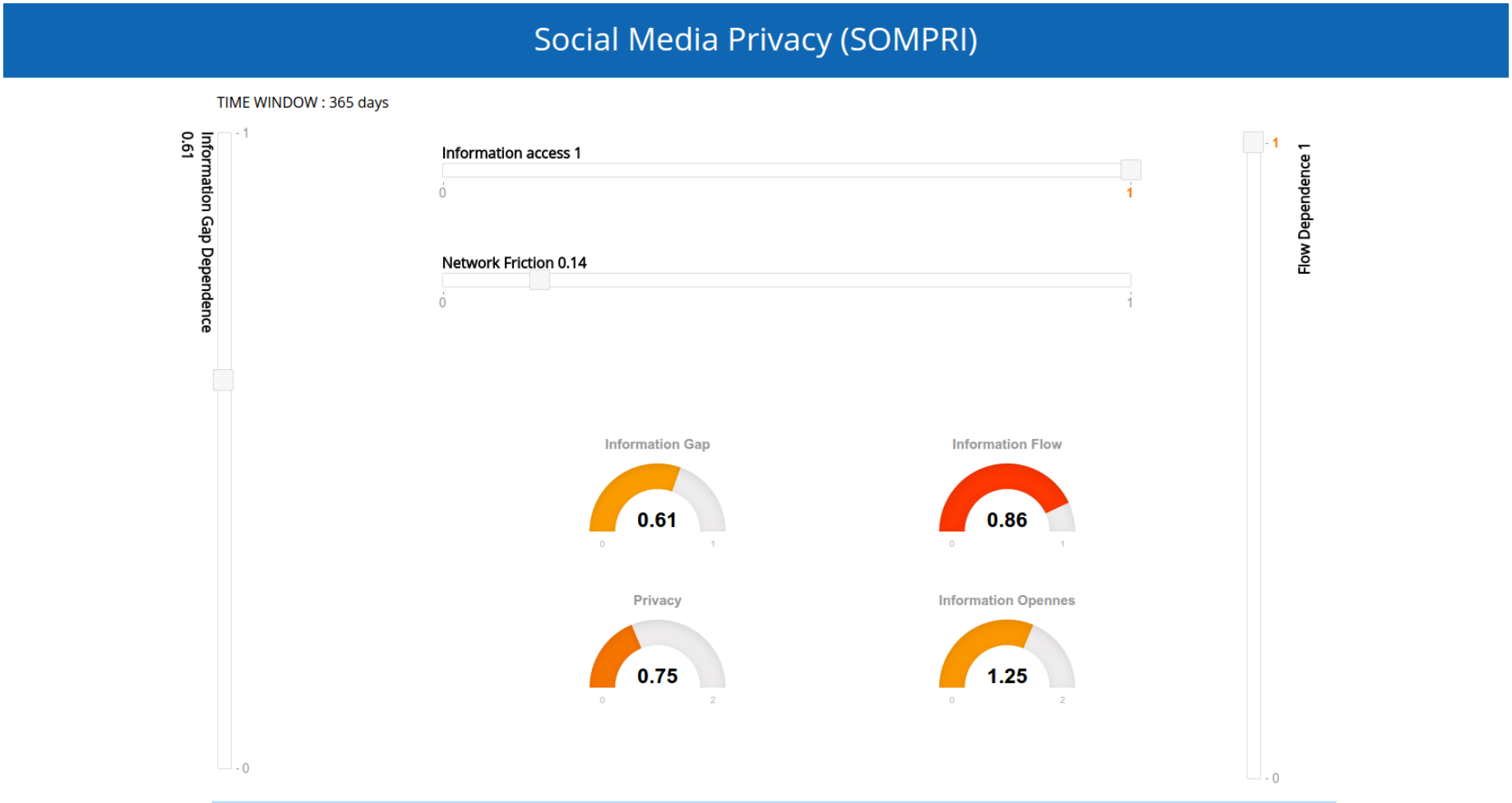


Figure: Facebook plug-in: results for User 3.

Comparison on data

The data for the three users is summarized in Table 1.

	IA	IG	NF	IF	IP
User1	0.23	0.32	1.00	0.00	1.32 (+0.32)
User2	0.47	0.49	0.20	0.80	0.69 (-0.31)
User3	1.00	0.61	0.14	0.86	0.75 (-0.25)

Table: Comparison of the Profile Cases.

Objectives for the experimental analysis

- Does the implementation 'faithfully' reflects (some aspects) of reality?
- Is the data indicative?
- Is the data reliable?
- Can we improve/Obtain different results with more data in the BN? (e.g. extension to third party apps)

A comparative analysis

Formal:

- **Technical View on Value**: defines the notion (in our case, in the context of information)
- general properties
- verified results
- required for analysis

Experimental:

- **Empirical View on Value**: measures the realization (in our case, code implements behaviour)
- network properties
- size analysis
- complexity analysis

Different Methodology, Different Experimental Setting

The context of SN characterizes our model by very different parameters (when e.g. compared to the trust model)

- Data set used: dynamic, changeable, at most correlations are found
- Data dependency: which data? which weights? which relations?
- Experiments: designed over tests performed on types of agents, results generalized (many limits)

Issues

- match user's values and the platform's API possibilities;
- identify data points to translate values in engineering elements;
- weight prior and posterior values, and define their dependence;
- analyse the effect of user's behaviour on the interpretation of probabilities;
- highlight the resulting limited and un-detailed knowledge;
- define conditions of approximation and the error bounds.

Sum up

- Methodologies in CS are different, but complementary
- In design with values (like privacy), this can be exploited to reach distinct objectives
- A model of privacy for has been considered first formally (for information systems) and then experimentally (for SNs)
- It highlights severe problems with the implementation of the scientific method in CS/SE

Thanks! (More) Questions?

References I


Balebako, R. and Cranor, L. F. (2014).
Improving App Privacy: Nudging App Developers to Protect User Privacy.
IEEE Security & Privacy, 12(4):55–58.


Floridi, L. (2005).
The ontological interpretation of informational privacy.
Ethics and Information Technology, 7(4):185–200.


Floridi, L. (2006).
Four challenges for a theory of informational privacy.
Ethics and Information technology, 8(3):109–119.

References II

Gross, R. and Acquisti, A. (2005).
Information Revelation and Privacy in Online Social Networks.
In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, WPES '05, pages 71–80, New York, NY, USA. ACM.

Gürses, S. F., Rizk, R., and Günther, O. (2008).
Privacy Design in Online Social Networks: Learning from Privacy Breaches and Community Feedback.
In *Proceedings of the International Conference on Information Systems, ICIS 2008, Paris, France, December 14-17, 2008*, page 90. Association for Information Systems.

Hazari, S. and Brown, C. (2014).
An Empirical Investigation of Privacy Awareness and Concerns on Social Networking Sites.
Journal of Information Privacy and Security, 9(4):31–51.

References III


Houghton, D. J. and Joinson, A. N. (2010).
Privacy, Social Network Sites, and Social Relations.
Journal of Technology in Human Services, 28(1-2):74–94.


Irani, D., Webb, S., Pu, C., and Li, K. (2011).
Modeling Unintended Personal-Information Leakage from Multiple Online Social Networks.
Internet Computing, IEEE, 15(3):13–19.


Jeonga, Y. and Coylea, E. (2014).
What Are You Worrying About on Facebook and Twitter? An Empirical Investigation of Young Social Network Site Users' Privacy Perceptions and Behaviors.
Journal of Interactive Advertising, 14(2):51–59.

References IV


Kang, T. and Kagal, L. (2010).
Enabling Privacy-Awareness in Social Networks.
In *Intelligent Information Privacy Management, Papers from the 2010 AAAI Spring Symposium, Technical Report SS-10-05, Stanford, California, USA, March 22-24, 2010*. AAAI.


Kelley, P. G., Cranor, L. F., and Sadeh, N. (2013).
Privacy As Part of the App Decision-making Process.
In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3393–3402, New York, NY, USA. ACM.


Krishnamurthy, B. and Wills, C. E. (2008).
Characterizing Privacy in Online Social Networks.
In *Proceedings of the First Workshop on Online Social Networks*, WOSN '08, pages 37–42, New York, NY, USA. ACM.

References V

Li, Y. (2012).
Theories in Online Information Privacy Research: A Critical Review and an Integrated Framework.
Decis. Support Syst., 54(1):471–481.

Li, Y., Li, Y., Yan, Q., and Deng, R. H. (2015).
Privacy Leakage Analysis in Online Social Networks.
Comput. Secur., 49(C):239–254.

Omoronyia, I., Cavallaro, L., Salehie, M., Pasquale, L., and Nuseibeh, B. (2013).
Engineering Adaptive Privacy: On the Role of Privacy Awareness Requirements.
In *Proceedings of the 2013 International Conference on Software Engineering*, ICSE '13, pages 632–641, Piscataway, NJ, USA. IEEE Press.

References VI


Takano, Y., Ohta, S., Takahashi, T., Ando, R., and Inoue, T. (2014).
MindYourPrivacy: Design and implementation of a visualization system for third-party Web tracking.
In *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*, pages 48–56.


Tan, X., Qin, L., Kim, Y., and Hsu, J. (2012).
Impact of privacy concern in social networking web sites.
Internet Research, 22(2):211–233.


Troncoso, C. (2013).
Bayesian inference to evaluate information leakage in complex scenarios.
In Puech, W., Chaumont, M., Dittmann, J., and Campisi, P., editors, *ACM Information Hiding and Multimedia Security Workshop*,

References VII

IH&MMSec '13, Montpellier, France, June 17-19, 2013, pages 1–2. ACM.

Warnier, M., Dechesne, F., and Brazier, F. (2015).
Design for the Value of Privacy.
In *Handbook of Ethics, Values and Technological Design*, pages 431–445. Springer.

Xu, W., Zhou, X., and Li, L. (2008).
Inferring privacy information via social relations.
In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 525–530.