ISSN 2186-7437

NII Shonan Meeting Report

No. 2013-5

Privacy by Transparency for Data-Centric Services

Prof. Dr. Isao Echizen
Prof. Dr. Günter Müller
Prof. Dr. Ryoichi Sasaki
Prof. Dr. A Min Tjoa

August 6–8, 2013

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Privacy by Transparency for Data-Centric Services

Organizers:
Isao Echizen (National Institute of Informatics, Japan)
Günter Müller (Albert-Ludwigs Universität Freiburg, Germany)
Ryoichi Sasaki (Tokyo Denki University, Japan)
A Min Tjoa (Vienna University of Technology, Austria)

August 6–8, 2013

1 Transparency and Relation to Privacy

The objective of the seminar was to investigate transparency as an extension or alternative to presently used mechanisms to ensure privacy. The reason is the popularity of data-centric services collecting personal data where users are mostly unaware of what private data are collected. Prevention as a basic principle, which underlies all proposed privacy mechanisms does not seem to be a solution. At the first glance, transparency seems to be a contradiction to privacy. At a second glance, transparency means access of a user to his or her data. The objective of transparency is the possibility to request a copy or deletion of data by an individual user. This simple definition however is technically difficult to realize and has societal consequences, if implemented without modification.

2 Scope of Seminar

This Shonan Seminar has addressed the following questions in selected presentations, panels, and discussions as well as individual statements:

A *Is Privacy and transparency a contradiction, since it cannot prevent violations?*

B *What is the relation between users privacy concerns and their trust in a particular service, based on the research available to this end?*

C *Does the available body of evidence support the assumption that more transparency would lead to more trust?*

D *Which transparency enhancing tools are available and in use at this moment, and what is experience with regard to enhance privacy?*

E *What are the societal requirements, and does privacy lead to behavioral changes, economic inefficiencies caused by a setback in technical progress?*

The talks [see section 6 of this report] have been classified and solicited according to these questions.

3 Privacy Enhancing Technology and Transparency Enhancing Technology

The evolution of privacy and security mechanisms occurred in distinguishable steps where access control composed of authentication and authorization is the unchallenged model for all privacy mechanisms. The developed mechanisms, of those most important are digital signatures, public key Infrastructures, and identity management [8, 18, 19], are called Privacy Enhancing Technology (PET). While anonymization [3] totally omits data for authentication, and Secure Multiparty Computing (SMC) [4] is a negotiation model, access control is control of provisions to grant access or deny access to data. In 2004 Park and Sandhu [13] extended the control level from the point of access to data usage to check obligations agreed upon when access to data was requested. The mechanisms encompass many privacy policy languages such as P3P [18] including its Freiburg variant ExPDT [14] allowing to compare policies. Sticky policies [8], secure logging [1], and data provenance [6] are examples of Transparency Enhancing Technology (TET) mechanisms. TET adds detection of violation while PET prevents violation if possible.

TET consists of signaling and screening functions. While signaling allows the specification of privacy rules, screening encompasses all mechanisms to control the enforcement and detection of the signaled rules. The components of TET are shown in the Figure 1.

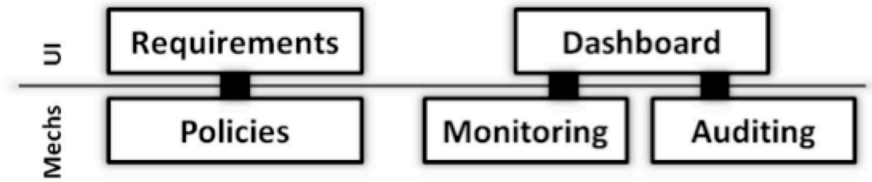

Figure 1: Mechanisms and User Interface.

Dashboards belong - like the requirements - to the user interfaces, while policies, monitoring and auditing are mechanisms. Dashboards, as offered in todays online social networks (OSN), experience a significant lack of trust by users originating from a lack of transparency.

4 Usage of Transparency Enhancing Technology

Usage of TET is an unsolved issue. The problem is described by the term data provenance, which consists of distribution and tracking of data access and usage to assure detection of violations [6]. Scalability during distribution and impossibility of tracking limits at present the use of TET. The so-called Freiburg

Transparency Meta Model - as shown in Figure 2 - is one proposal for a framework to deduce needed requirements [10]:

- The *user or business layer* defines the data objects, business processes, the assets of people and companies, as well as the privacy and security guidelines to be followed.
- The *application layer* contains the IT services, data schemes, and mechanisms, which are required for the usage enforcement of data after PET has been applied for the provision phase.
- The *infrastructure layer* provides the software and hardware needed to automate the execution of security and privacy.

The Meta Model can be split in the respective layers according to the time points they act upon in design-time, run-time, and audit-time each having clearly separable mechanisms to enforce privacy.

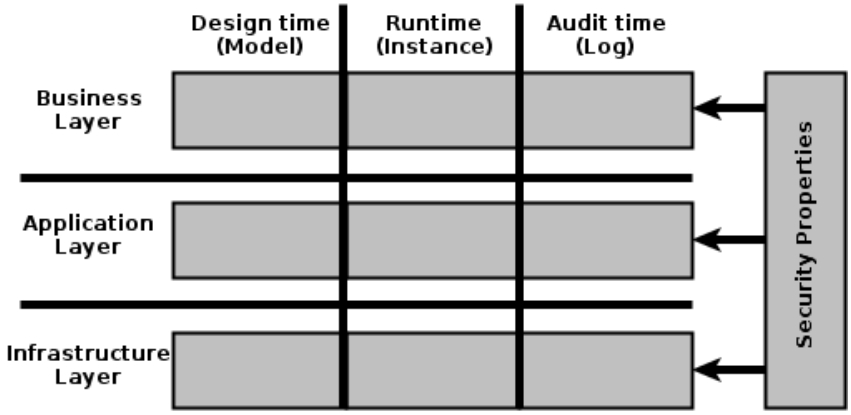

Figure 2: Freiburg Privacy Model.

5 Big Data and Transparency

With the advent of Big Data analytics of the data for improved decision making and profiling, people adds a new dimension to privacy. No legal body about data protection covers inferences, even though their existence depends upon private data [12]. Privacy of humans is intrinsically an individual and normative activity as it aims to either maintain a desirable state or adapt and transform towards a more desirable state. Inferences are properties of aggregates, not directly about individuals. As shown in Figure [**?**] the usage of data sources - so far not in focus of IT - poses a new challenge to transparency and privacy, since tracing back an event to the processing of an individual data item is almost impossible.

Privacy and Big Data may lead to the emergence of unintended, unpredictable safety, reliability, and acceptance problems [12], since the privacy exposure is the analytical capability as derived from [9]:

1. **Volume:** As of 2012 about 2.5 exabytes are generated every day, and this is doubling every forty months. The interesting part is that the data have mostly not been collected so far, is new, of low density and until now it has not been in the focus of Information Technology (IT).

2. **Velocity:** Cyper-Physical Systems (CPS) speed is more important than volume. Real time or close to real-time information as is expected in pervasive computing makes it possible to be more agile on users behavior.

3. **Variety:** Big data draws patterns from all sorts of structured and unstructured formats including textual messages, audit data, or images. Data is received from sensors or GPS signals, from cell phones, or gas stations when a digitized form of payment is used. Many forms of data collection are new, and may not have been in the focus of classical IT.

4. **Analytics and inferences:** Big Data and privacy is more about analytics than it is about storage. Connecting to new data sources, detecting correlations and behavioral patterns and giving the results, as an input to improved decision-making will be the source of data for new services, and are called inferences [5].

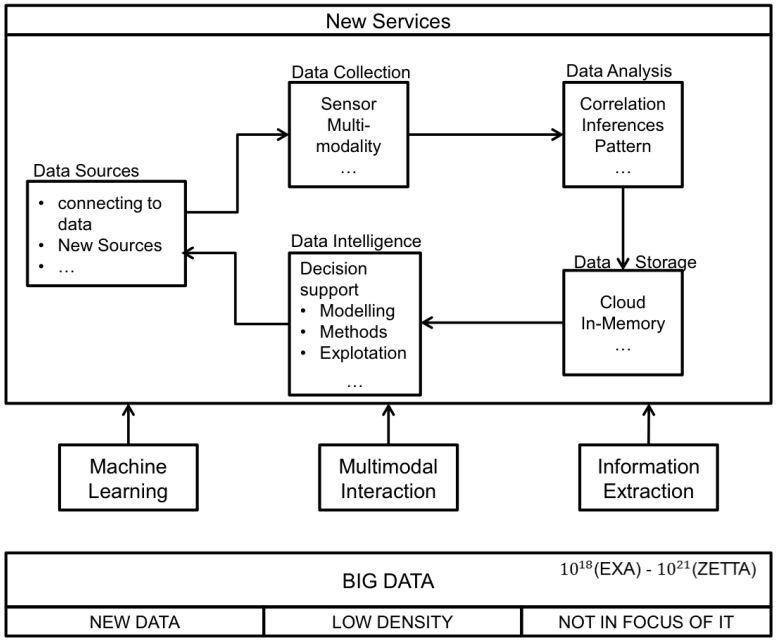

Figure 3: Big Data and Privacy [10].

6 Right to Be Forgotten?

While PETs is a set of mechanisms protecting privacy by eliminating or minimizing personal data, transparency seems to contradict privacy. TET and its underlying concepts of transparency were defined as insight in how users data is being collected, stored, processed and disclosed. TET is viewed as set of tools providing this insight, and at the same time allows exercising a possible correction of proven privacy violations. This property of transparency is a danger for privacy in itself. The right to forget or exercise changes on data has its limits in the right of future generations to have a correct image of a certain time about a set of events. It is also a limit to the right to forget, if the actions of one person has impact on others. TET mechanisms may be misused to request the deletion of unpleasant but true facts.

References

[1] R. Accorsi: A secure log architecture to support remote auditing. Mathematical and Computer Modeling, Elsevier, doi:10.1016/j.mcm.2012.06.35, 2012.

[2] D. Basin, M. Harvan, F. Klaedtke, and E. Zalinescu: MONPOLY: Monitoring Usage-Control Policies. RV, 360–364, 2011.

[3] D. Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM 24(2), ACM, 84–88, 1981.

[4] D. Dolev and A.C. Yao: On the Security of Public Key Protocols. IEEE Transactions on Information Theory 2(29), IEEE Press, 198–208, 1983.

[5] J. Epstein, Security Lessons Learned from Socit Gnrale. IEEE Security & Privacy 6(3): 80-82 (2008).

[6] S. Haas, S. Wohlgemuth, I. Echizen, N. Sonehara, and G. Müller: Aspects of Privacy for Electronic Health Records. Int. Journal of Medical Informatics, Special Issue: Security in Health Information Systems 80(2), Elsevier, e26–e31, 2011.

[7] L. Kagal and H. Abelson: Access Control is an Inadequate Framework for Privacy Protection. W3C Workshop on Privacy for Advanced Web APIs, 2010. Available at http://www.w3.org/2010/api-privacy-ws/papers/privacy-ws-23.pdf

[8] G. Karjoth, M. Schunter, and M. Waidner: Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data. 2nd Workshop on Privacy Enhancing Technologies. LNCS 2482, Springer, 69–84, 2003.

[9] A. McAfee and E. Brynjolfsson: Big Data: The Management Revolution, Harvard Business Review, pp. 59-68, October, 2012.

[10] G. Müller and R. Accorsi: Why are Business Processes not secure? In: M. Fischlin and St. Katzenbeisser: Festschrift in Honor of Johannes Buchmann, Lecture Notes in Computer Science, Springer 2013. (in print)

[11] G. Müller and K. Rannenberg (eds.): Multilateral Security in Communications - Technology, Infrastructure, Economy. Addison-Wesley, 1999.

[12] G. Müller and W. Wahlster: Putting Humans back in the feedback loop of social infrastructures, In: Informatik Spektrum, 2013. (in print)

[13] J. Park and R. Sandhu: The $UCON_{ABC}$ Usage Control Model. 24th ACM Transactions on Information and System Security 7(1), ACM, 128–174, 2004.

[14] S. Sackmann and M. Kähmer: ExPDT: A Policy-based Approach for Automating Compliance. Wirtschaftsinformatik 50(5), Gabler, 366–374, 2008.

[15] A. Schröpfer, F. Kerschbaum, and G. Müller: L1 - An Intermediate Language for Mixed-Protocol Secure Computation, COMPSAC 11 Proceedings of the 2011 IEEE 35th Annual Computer Software and Applications Conference, IEEE Press, 298–307, 2011.

[16] N. Sonehara, I. Echizen, S. Wohlgemuth, G. Müller, and A. Tjoa (eds.): Proceedings of the International Workshop on Information Systems for Social Innovations (ISSI) 2009, http://www.nii.ac.jp/issi, National Center for Sciences, 2009.

[17] K. Takaragi, R. Sasaki, and S. Singai: A Probability Bounds Estimation Method in Markov Reliability Analysis. IEEE Transactions on Reliability 3(35), IEEE Press, 257–261, 1985.

[18] R. Wenning and M. Schunter (eds.): The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. W3C Working Group Note 13, 2006. Available at http://www.w3.org/TR/P3P11/.

[19] S. Wohlgemuth and G. Müller: Privacy with Delegation of Rights by Identity Management. In: *ETRICS 2006*, LNCS Vol. 3995; Springer, pp. 75–190, 2006.

Overview of Talks

The remainder of this report gives the outlines of the talks and contributions ordered according to the above questions.

Ad A: Scope of Problem Present and Future Privacy Challenges

Is Transparency a helpful Paradigm to enforce Privacy?

Müller, Günter (Albert-Ludwigs Universität Freiburg, Germany)

Some - even superficial - analysis of the impact of privacy mechanisms show that actually none of them is used, to really enforce privacy. This is despite the fact that the majority of users find privacy a key issue. This seemingly paradoxical behavior calls for a rethinking of the fundamental concepts of privacy enhancing technologies or just give up unreasonable demands. This motivational talk proposes to complement the prevention paradigm by detective mechanisms. Privacy as understood in IT is shown as a follower of technical progress, and it is argued that the business interest and contributions to productivity and welfare is based upon the availability of data, which may be the reason why privacy becomes a revised specification. The data centric view however, has in addition to the interests of service providers an exposure for privacy that may lead to inefficient societies. The main issue is seen in the enforcement of both PET and TET alike. As a Meta Model the Freiburg framework is proposed.

Current Developments in Public Policy for Privacy and Transparency: Implications for the Proposed Right to Be Forgotten

Longstaff, Patricia Hirl (Syracuse University, USA)

I will give an overview of current debates in the US on these public policy issues and the supporters and opponents of the proposals.

Policy Switching in Emergency

Maruyama, Hiroshi (The Institute of Statistical Mathematics, Japan)

I will argue that the priorities on security/privacy requirements change in the face of emergency such as a natural disaster. We are working on a security architecture to incorporate policy switching, and present some of the insights and challenges obtained from the work.

Ad B: Privacy Concerns and Privacy Support

Human Resource-Centric Services: Privacy and Cost

Akiyoshi, Masanori (Hiroshima Institute of Technology, Japan)

Recent crowd sourcing services make it possible to provide anonymous bridges between service requester and service provider. For instance, translation from

one language to the other is one good example service on the internet. However this service seems to have negative aspect in addition to the flexible aspect. Since the data is transferred from service requester to anonymous service provider, it is significant to guarantee privacy on requester and his/her data. Of course some regulations are set in advance to establish such services, however, anonymity still makes negative effect on it. As spam workers or poor level workers in crowd sourcing services are identified, workers on privacy violation likelihood should be identified. Then we can imagine that such identification costs much along with the size of crowd sourcing. I would like to discuss such emerging issues to be tackled hereafter on human resource-centric services.

Usable Abstractions for Policy Authoring by Non-Experts

Bauer, Lujo (Carnegie Mellon University, USA)

Key aspects of transparency in protecting private data from unauthorized use include supporting intuitive methods for users to express their desired privacy preferences and to get feedback about already implemented preferences and their effects on data access. I will discuss several efforts to make progress in this space, including: two new types of user interfaces for policy authoring, which we call expandable grids and proximity displays; a just-in-time approach for allowing users to specify policies, which we call reactive policy creation; and initial effort to allow users to specify policies via metadata tags.

Method for Preventing Privacy Invasion through Face Recognition from Camera Images

Echizen, Isao (National Institute of Informatics, Japan)

A method is proposed for preventing unauthorized face image revelation through unintentional capture of facial images. Methods such as covering the face and painting particular patterns on the face effectively prevent detection of facial images but hinder face-to-face communication. The proposed method overcomes this problem through the use of a device worn on the face that transmits near-infrared signals that are picked up by camera image sensors, which makes faces in captured images undetectable. The device is similar in appearance to a pair of eyeglasses, and the signals cannot be seen by the human eye, so face-to-face communication is not hindered. Testing of a prototype privacy visor showed that captured facial images are sufficiently corrupted to prevent unauthorized face image revelation by face detection.

Can Linked Open Data be Used to Empower Continuous Auditing?

Tjoa, A Min (Competence Center for Excellent Technologies - Secure Business Austria & Vienna University of Technology, Austria)

The scattered information on the Web can form a global data graph that connects distributed resources and facilitates the discovery of new resources. In this context Linked Data introduces some simple and effective principles for

publishing and connecting structured data on the Web. Linked Data has gained momentum among governments, in the academic and business world, and in the public sector over the last few years. Today a growing number of high quality and public Linked Data resources are published on the Web which can benefit the decision makers and authorities at the national and international levels to overcome the data gaps and improve the information availability.

In this talk, the recent advancements of Linked Data and Linked Open Data (LOD) approaches for capturing, managing, and distribution of information will be explored and their potential for addressing Continuous Auditing requirements will be highlighted.

Ad C: Privacy and Trust

Empowering Patients (Citizens) to Information Self-Determination

Katt, Basel (University of Innsbruck, Austria)

One of the main aspects of privacy protection is the users right of information self-determination. In the area of shared electronic health records, legal requirements in some countries, like Austria, require citizens to be enabled to decide who should be allowed to use what medical data and in which way. In this talk, ScenBAC (Scenario based Access Control), an access control administration model, will be presented. ScenBAC enables non-security stakeholders (e.g., citizens) to define own privacy policies.

Effect of External Information on Anonymity and Transparency

Yoshiura, Hiroshi (University of Electro-Communications, Tokyo, Japan

We show the limits of social network anonymity based on several case studies including our development and evaluation of Ineluctable Background Checking System on Social Networks. We then discuss in general terms the limits of anonymity when an attacker can use external information that an anonymizer cannot know in advance. The complimentary role of transparency is discussed to mitigate this problem and to protect anonymized personal information from de-anonymization while enabling flexible use of it.

Ad D: Present Transparency Enhancing Tools

On the Limits of Transparency: Why it is not always enforceable

Kerschbaum, Florian (SAP Applied Research, Germany)

I will give some examples, such as insider threats in banks and tax cases, and argue why privacy is a conflict of interest that needs clear and verifiable settlement rules.

Google Dashboard: Transparency Way to more Privacy?

Flatscher, Rony G. (WU Vienna, Austria)

The presentation will analyze the functionality of Google dashboard with regard to improving transparency for Privacy. It is intended to demonstrate some scenarios with to show accomplishments and deficits.

Usage Control and Sticky Policies in Practice

Lotz, Volkmar (SAP Research, France)

This talk focuses on pragmatic aspects of enforcing privacy controls via sticky policies. These aspects include policy language design, user-defined policies, system architecture, migration strategy, scope of control, and performance concerns for industrial scale applications. We present an implementation of a sticky policy engine that takes advantage of in memory databases to scale performance.

Formal Models of (Privacy) Requirements for Off- and Online Validation

Padget, Julian (University of Bath, UK) and Satoh, Ken (National Institute of Informatics, Japan)

Our focus is on the specification, validation and evolution of policy to meet policy makers and users requirements. We will describe a simple action language for the capture of policy requirements and show how to construct a corresponding model by means of Answer Set Programming. By using an answer set solver, it is possible to explore all possible traces for all the actors for all event orderings, which permits the designer to check whether desired global properties are upheld or not. The same model may also be used check compliance in live systems or advise participants whether an action or actions are policy compliant. We look forward to seeing what policy modelling challenges the seminar identifies and how these fit our approach. While this provides a technology for evaluating policy, it leaves unaddressed the tricky issues of policy capture - is it desirable to employ specialist policy designers, or somehow to capture policy through practice? - and accessibility of policy language, while our approach depends on first-order logic, users may prefer to author and read specifications in natural language or some diagrammatic notation.

Transparency Enhancing Technologies and the Inherent Intransparency of Agile Big Data Mining

Rannenberg, Kai (Goethe University Frankfurt, Germany)

Transparency Enhancing Technologies (TETs) are devised as a paradigm to support privacy protection in data-rich services by enabling users to understand the potential processing of their data as a basis for making decisions about data flows and data processing.

This presentation will reflect on TETs based on early experiences from the projects PRIME (Privacy and Identity Management for Europe) and FIDIS (Future of Identity in the information society).

The presentation will start with an introduction into a successful TET for protecting users in Telco operated location based services, which will also explain 4 success factors for trust enabling transparency: Transparency of the infrastructure, Transparency of the operations, Transparency of the options, and Transparency assurance.

Then the presentation will analyze the profiling functionality of agile big data mining and the related requirements for privacy protection via automated TETs. This will explain the impact on the relevant transparency success factors including the spiral of automated transparency assurance. Time permitting some ideas on overcoming the situation will be sketched.

Proposal and Evaluation of an Evidence Preservation Method for Use in a Common Number System

Sasaki, Ryoichi (Tokyo Denki University, Japan)

In recent years, the introduction of a common number system has been planned by the government of Japan, and the possibility of illegal use of personal information in this system has been considered. Access records in log files are analyzed when illegal use is being investigated. Therefore, a method by which to maintain the reliability of the log file becomes a significant problem, because alteration of digital data is very easy. Moreover, in the case of a common number system, a method that keeps verification nature as well as secret from a certain organization is needed, because various organizations will have access to the common number system. In the present paper, we propose an evidence preservation method that enables privacy and concealment to be maintained by introducing a cipher system and a hysteresis signature.

Dynamic Pseudonym Scheme for Improving the Utility of Location Data Sets

Minami, Kazuhiro (The Institute of Statistical Mathematics, Japan)

Anonymization is a primary way to make a data set containing private information available in the public so that people who is interested in that data set can perform any analytic analysis. However, when we anonymize a location data set, the utility of that data set is significantly degraded since all the trajectory information of mobile users in the set is lost. We present a dynamic pseudonym scheme that takes a better balance between data utility and privacy protection of a location data set and argue that the proposed scheme makes our society more transparent by improving the utility of data sets available in the public.

Ad E: Non Functional Methods and Behavioral Impacts

Finding an Approach to Data Protection including Privacy Protection from the Perspective of Japanese Criminal Law

Nishigai, Yoshiaki (The University of Tokyo, Law and Politics, Japan)

My interest is the data protection of Japanese criminal law. Privacy protection is, I think, one of the fields of informational law. It could be said that Japanese criminal law does not have sufficient regulation especially in the WWW. In the case of consideration of privacy protection from the view point of law, it is indispensable to know the technology about a way of data protection or other technological data control ways in order to find (or think) better legal interpretation.

Data Owner Controlled Access Control Mechanisms for Healthcare Applications

Schrittwieser, Sebastian (SBA Research, Austria)

The vast majority of data security and access control mechanisms in healthcare systems are centrally controlled by administrators who are a major threat to the patients privacy. Apart from administrators, other internal persons, such as hospital staff members, may exploit their access rights to snoop around in private health data. In this talk I want to present a security protocol for data privacy that is strictly controlled by the data owner. It integrates pseudonymization and encryption to create a methodology that uses pseudonyms as access control mechanism. Further, I want to discuss the impact of the introduced transparent access controls on the users trust in the system.

Transparency meets Win-Win Relationships by Privacy Requirements Engineering

Yoshioka, Nobukazu (National Institute of Informatics, Japan)

Service providers require to use the privacy data to provide Service Providers require to use the privacy data to provide tailor-made functionally, such as recommendation, for users. In other hand, users have rights to preserve their privacy, so they can decide their privacy policy when they use services. If users want to provide their privacy data because of afraid with a service, they cannot have value of the service. So, it is important to get win-win relationships between service providers and the users. Transparency plays a key role for for the relationships because it is a kind of communications with users. In my talk, I describe how to identify win-win relationships with transparency from the Requirements Engineering (RE) point of view. RE provides not only methods to specify system requirements but also identify win-win relationships between stakeholders to decide the best solution for them. Especially, we can identify why privacy data are needed and the users value with Goal-Oriented Requirements Engineering. We, therefore, can know how to realize good transparency of a service. My talk will include research issues on software engineering to provide privacy transparency of services.

Panel Discussion: Is Transparency the new Privacy Paradigm?

Coordinator: Kai Rannenberg
Discussants: Ryoichi Sasaki, Isao Echizen, Hiroshi Yoshiura, Volkmar Lotz

Transparency does not give privacy, but allows to detect misuse. One needs to explicitly define what privacy attempts to achieve separately. It is felt to be necessary to reason about the meaning of privacy, which probably is perceived differently by younger generations. Transparency will be a general means of controlling, achieving accountability, turning black-box principles (secrets) into white-boxes that can be studied and analyzed. Transparency is generally an important principle for democracy. Some statements from the audience and the panel may give the scope of discussion: One conclusion the participants agreed upon is that privacy is for the weak (Kai Rannenberg) and transparency is for the powerful (A Min Tjoa), transparency is option of the weaker to learn what the powerful do (Günter Müller). Transparency is a means to balance asymmetry (Isao Echizen). De-anonymization is easy to achieve with advent of public data that transparency is possible means to maintain privacy (Hiroshi Yoshiura).

Participants

Akiyoshi, Masanori, Hiroshima Institute of Technology, Japan
Bauer, Lujo, Carnegie Mellon University, USA
Echizen, Isao, National Institute of Informatics, Japan
Flatscher, Rony G., WU Vienna, Austria
Katt, Basel, University of Innsbruck, Austria
Kerschbaum, Florian, SAP Applied Research, Germany
Longstaff, Patricia Hirl, Syracuse University, USA
Lotz, Volkmar, SAP Research, France
Maruyama, Hiroshi, The Institute of Statistical Mathematics, Japan
Minami, Kazuhiro, The Institute of Statistical Mathematics, Japan
Müller, Günter, Albert-Ludwigs Universität Freiburg, Germany
Nishigai, Yoshiaki, The University of Tokyo, Japan
Padget, Julian, University of Bath, UK
Rannenberg, Kai, Goethe Universität Frankfurt, Germany
Sasaki, Ryoichi, Tokyo Denki University, Japan
Satoh, Ken, National Institute of Informatics, Japan
Schrittwieser, Sebastian, SBA Research, Austria
Shareeful, Islam, University of East London, UK
Tjoa, A Min, Competence Center for Excellent Technologies - Secure Business Austria & Vienna University of Technology, Austria
Yoshioka, Nobukazu, National Institute of Informatics, Japan
Yoshiura, Hiroshi, University of Electro-Communications, Tokyo, Japan