Themenliste des Proseminars im Sommersemester 2015

Dr. Michael Schiffers Dr. Nils gentschen Felde Prof. Dr. Dieter Kranzlmüller

I. Allgemeine Hinweise

Für alle Themen gelten als wünschenswerte Ergebnisse mindestens die folgenden: 1. Erklären Sie das Problem! 2. Ordnen Sie das Thema / die Konzepte! 3. Stellen Sie Ergebnisse vor und betrachten Sie diese kritisch! Der kritische Umgang mit wissenschaftlicher Literatur ist ein wesentliches Ziel des Seminars!

Die jeweiligen Literaturhinweise betrachten Sie bitte als Vorschläge und Einstiege in das Thema. Sie sind manchmal zu weit gefasst. Es ist ausdrücklich erwünscht, dass Sie selbständig recherchieren und bibliographieren. Sie sollen also keine „erweiterte Inhaltsangabe“ oder „Nacherzählung“ der angegebenen Literatur abliefern. Besprechen Sie die Vorgehensweise mit Ihrem Betreuer.

II. Themen

A. Positioning

1 **Privacy versus Security**

a) Inhalt:
Privacy and security often tend to conflate. However, security and privacy can, and should, be treated as distinct concerns. Privacy discourse involves difficult normative and technical decisions about competing claims to legitimate access to, use of, and alteration of personal information. Security implements those choices – it mediates between information and privacy selections as it preserves confidentiality, integrity and availability of information; in addition, other properties such as authenticity, accountability, non-repudiation and reliability can also be involved. However, this is only one aspect of security. The other relates to the more political aspect of societal security. In this presentation we will *not* focus on this latter aspect.

b) Literatur:

- Bambauer: Privacy versus Security
- Avizienis et al.: Basic Concepts and Taxonomy of Dependable and Secure Computing
- Boran: IT Security Cookbook (Chapter 7)
- Cavoukian, Dixon: Privacy and Security by Design: An Enterprise Architecture Approach

c) Betreuer: Stefan Metzger (LRZ)

d) Bearbeiter: noch offen

B. Formal Basics and Quantification

2 **k-Anonymity**

a) Inhalt:
k-Anonymity is a property possessed by certain anonymized data to solve the problem: „Given person-specific field-structured data, produce a release of the data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful“. l-Diversity and t-closeness address shortages of the k-anonymity model.

b) Literatur:

- Sweeney: k-Anonymity: A Model for Protecting Privacy
- El Emam, Kamal Dankar: Protecting Privacy Using k-Anonymity
- Machanavajjhala et al.: l-Diversity: Privacy Beyond k-Anonymity
- Li et al.: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity

c) Betreuer: Michael Schiffers

d) Bearbeiter: noch offen

3 **Differential Privacy**

a) Inhalt:
Differential privacy is a definition of a privacy goal tailored to privacy-preserving data analysis. It provides a mathematically rigorous theory of privacy. Although it is the strongest notion of privacy known to date, it is also known that no deterministic algorithm can guarantee differential privacy.

b) Literatur:

- Dwork: Differential Privacy: A Survey of Results
- Dwork: A Firm Foundation for Private Data Analysis
- Dwork et al.: Calibrating Noise to Sensitivity in Private Data Analysis
- Dwork et al.: Differential Privacy: A Primer for the Perplexed

c) Betreuer: Michael Schiffers

d) Bearbeiter: noch offen

4 **Measuring Privacy**

a) Inhalt:

Approaches to measure privacy concerns are fragmented and often ad-hoc, at the detriment of reliable results. The need for measurement instruments for privacy concern is twofold. First, attitudes and opinions about data protection cannot be established and compared without reliable mechanisms. Second, behavioural studies, notably in technology acceptance and the behavioural economics of privacy require measures for concern as a moderating factor.

b) Literatur:

- Preibusch: Guide to measuring privacy concern: Review of survey and observational instruments
- Bertino et al.: A Survey of Quantification of Privacy Preserving Data Mining Algorithms
- Braunstein et al.: Indirect Content Privacy Surveys: Measuring Privacy Without Asking About It
- Wang: A Privacy Measure of Online Personally Identifiable Information

c) Betreuer: Matthias Maiterth

d) Bearbeiter: noch offen

---

5 **De-Anonymization**

a) Inhalt:

Any information that distinguishes one person from another can be used for re-identifying anonymous data. Re-identification algorithms are agnostic to the semantics of the data elements. The emergence of powerful re-identification algorithms demonstrates not just a flaw in a specific anonymization technique(s), but the fundamental inadequacy of the entire privacy protection paradigm based on „de-identifying“ data. De-identification provides only a weak form of privacy.

b) Literatur:

- Narayanan, Shmatikov: Robust De-Anonymization of Large Datasets
- Ohm: Broken promises of privacy: Responding to the surprising failure of anonymization
- Balduzzi et al.: Abusing Social Networks for Automated User Profiling
- Almishari et al.: Are 140 Characters Enough? A Large-Scale Linkability Study of Tweets

c) Betreuer: Karl Fürlinger

d) Bearbeiter: noch offen

C. Privacy and Infrastructure Aspects

6 **Privacy in Clouds**

a) Inhalt:

Cloud computing has generated significant interest in both academia and industry, but it's still an evolving paradigm. Essentially, it aims to consolidate the economic utility model with the evolutionary development of many existing approaches and computing technologies, including distributed services, applications, and information infrastructures consisting of pools of computers, networks, and storage resources. Confusion exists in IT communities about how a cloud differs from existing models and how these differences affect its adoption. While cloud computing could significantly enhance collaboration, agility, and scale, thus enabling a truly global computing model over the Internet infrastructure, without appropriate security and privacy solutions designed for clouds, this potentially revolutionizing computing paradigm could become a huge failure.

b) Literatur:

- Wei et al.: Security and privacy for storage and computation in cloud computing
- Kalloniatis et al.: Towards the design of secure and privacy-oriented Information Systems in the Cloud: Identifying the major concepts
- Neela, Saravanan: Privacy Preserving Approaches in Cloud: a Survey
- ITU-T: Privacy in Cloud Computing
- de Hert et al.: The new cloud computing ISO/IEC 27018 standard through the lens of the EU legislation on data protection
- Itani et al.: Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures

c) Betreuer: Bastian Kemmler (LRZ)

d) Bearbeiter: noch offen

---

7 **Location Privacy in Mobile Infrastructures**

a) Inhalt:

The problem of privacy in human mobility data shows that human mobility traces are highly identifiable with only a few spatio-temporal points. Even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual's privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals. Location obfuscation is a promising approach to protect the location-privacy of mobile users in location-based services.

b) Literatur:

- de Montjoye et al.: Unique in the Crowd: The privacy bounds of human mobility
- Song et al.: Not So Unique in the Crowd: a Simple and Effective Algorithm for Anonymizing Location Data
- Shokri et al.: Protecting Location Privacy: Optimal Strategy against Localization Attacks

c) Betreuer: Vitalian Danciu

d) Bearbeiter: noch offen

---

8 **Privacy and Big Data**

a) Inhalt:
One of the major applications of future generation parallel and distributed systems is in big-data analytics. Data repositories for such applications currently exceed exabytes and are rapidly increasing in size. Beyond their sheer magnitude, these datasets and associated applications' considerations pose significant challenges for privacy preserving.

b) Literatur:

- Kambatla et al.: Trends in big data analytics
- Aggarwal, Yu: A General Survey of Privacy-Preserving Data Mining Models and Algorithms
- Schadt: The changing privacy landscape in the era of big data
- Tene, Polonetsky: Big Data for All: Privacy and User Control in the Age of Analytics
- Executive Office of the President: Big Data and Privacy: A Technological Perspective
- International Working Group on Data Protection in Telecommunications: Working Paper on Big Data and Privacy

c) Betreuer: Nils gentschen Felde

d) Bearbeiter: noch offen

---

9 **Browser Tracking**

a) Inhalt:
Advanced web tracking mechanisms are hard to control, hard to detect and resilient to blocking or removing. There exist subtle pitfalls (such as failing to clear state on multiple browsers at once) in which a single lapse in judgment can shatter privacy defenses. This suggests that even sophisticated users face great difficulties in evading such web tracking techniques.

b) Literatur:

- Acar et al.: The Web Never Forgets: Persistent Tracking Mechanisms in the Wild
- Eckersley: How Unique is Your Browser?
- Mowery, Shacham: Pixel Perfect: Fingerprinting Canvas in HTML5
- Pan et al.: I Do Not Know What You Visited Last Summer – Protecting users from third-party web tracking with TrackingFree browser

c) Betreuer: Felix von Eye (LRZ)

d) Bearbeiter: noch offen

---

10 **Privacy and Intrusion Detection Systems**

a) Inhalt:
The relationship between intrusion detection and privacy is particularly interesting. In order to analyze the actions performed by users and applications, intrusion detections systems (IDS) need to collect and retain data about a user's behavior. In contrast, privacy is concerned with the right of individuals to determine if, when, how, and to what extend data about themselves will be collected, stored, transmitted, used, and shared with other. One approach to balancing these interests is replacing identifying features in audit data with pseudonyms and to reduce pseudonym linkability effects.

b) Literatur:

- Büschges: Privacy Enhanced Intrusion Detection
- Niksefat et al.: ZIDS – A Privacy-Preserving Intrusion Detection System using Secure Two-Party Computation Protocols
- Venter et al.: PIDS: A Privacy Intrusion Detection System
- gentschen Felde: Ein föderiertes Intrusion Detection System für Grids
- Manshaei et al.: Game Theory Meets Network Security and Privacy

c) Betreuer: Nils gentschen Felde

d) Bearbeiter: noch offen

D. Privacy-Preserving Techniques

11 **Fuzzy-Based Approaches**

a) Inhalt:
The primary goal of privacy preserving is to hide the sensitive data before it gets published. For example, a hospital may release patient's records to enable the researchers to study the characteristics of various diseases. The raw data contains some sensitive information of individuals, which must not be published to protect individual privacy. However, using other published attributes and additional external data allows retrieving personal identities. One approach to preserve sensitive information is using fuzzy logic.

b) Literatur:

- Karthikeyan et al.: A Fuzzy Based Approach for Privacy Preserving Clustering
- Sridhar, Babu: A Fuzzy Approach for Privacy Preserving in Data Mining
- Naga Lakshmi, Sandhya Rani, Babu: Privacy Preserving Clustering Based on Fuzzy Data Transformation Methods
- BSI: Study of the Privacy and Accuracy of the Fuzzy Commitment Scheme; BioKeyS III–Final Report

c) Betreuer: Felix von Eye (LRZ)

d) Bearbeiter: noch offen

---

12 **Immunology Inspired Approaches**

a) Inhalt:
Privacy protection issues usually raise from insufficient user privacy control mechanisms offered by service providers, unauthorized usage of user's data, and lack of appropriate privacy protection schemes for user's data at servers. Recently privacy protection models based on immunology inspired concepts have been proposed to provide automatic detection

and blocking of sensitive user information revealed in social communications and elsewhere. However, demands for data availability and the criteria for confidentiality are continually evolving, complicating the task of protecting sensitive data. Current encryption technology and query restriction help ensure confidentiality, but neither solution is appropriate for all applications. In the case of encryption, the ability to search data records is hindered; in the case of query restriction, individual records are vulnerable to insider attacks and their security can be compromised by tracker attacks.

b) Literatur:

- Lo, Yohan: Danger Theory-based Privacy Protection Model for Social Networks
- Esponda: Everything That is Not Important: Negative Databases
- Esponda et al.: Protecting data privacy through hard-to-reverse negative databases
- Esponda et al.: Negative representations of information
- Dasgupta, Niño: Immunological Computation: Theory and Applications (Chapters 4 and 7)

c) Betreuer: Michael Schiffers

d) Bearbeiter: noch offen

---

13 **The Onion Router (Tor)**

a) Inhalt:

Tor enables anonymous communication. Tor directs Internet traffic through a free, worldwide, volunteer network to conceal a user's location and usage from anyone conducting network surveillance or traffic analysis. Onion routing is implemented by encryption in the application layer of a communication protocol stack, nested like the layers of an onion, used to anonymize communication. Tor encrypts the original data, including the destination IP address, multiple times and sends it through a virtual circuit comprising successive, randomly selected Tor relays. Each relay decrypts a layer of encryption to reveal only the next relay in the circuit in order to pass the remaining encrypted data on to it. The final relay decrypts the innermost layer of encryption and sends the original data to its destination without revealing, or even knowing, the source IP address. An adversary unable to defeat the strong anonymity that Tor provides may try to de-anonymize the communication by other means.

b) Literatur:

- Dingledine et al.: Tor: The Second-Generation Onion Router
- Sun et al.: RAPTOR: Routing Attacks on Privacy in Tor
- Johnson et al.: Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries

c) Betreuer: Stefan Metzger (LRZ)

d) Bearbeiter: noch offen

---

14 **Privacy by Design**

a) Inhalt:

Complex and rapid technological change (e.g., emerging analytics) may create privacy harms as a by-product (e.g., more powerful analytics may inadvertently make it possible to re-identify individuals over large data sets). Ideally, privacy needs to be embedded, by default, during the architecture, design and construction of the processes. This is the central motivation for „Privacy by Design (PbD)“ which is aimed at reducing risks of privacy harm from arising in the first place. PbD is based on seven Foundational Principles. It emphasizes respect for user privacy and the need to embed privacy as a default condition, but preserves a commitment to functionality. This approach transforms individual privacy issues from a pure policy or compliance issue into a general imperative. PbD is focused on processes rather than a singular focus directing technical outcomes. Today, PbD is widely recognized internationally as the standard for developing privacy compliant information systems.

b) Literatur:

- Cavoukian: Privacy by Design: The 7 Foundational Principles: Implementation and Mapping of Fair Information Practices
- Gürses et al.: Engineering Privacy by Design
- zahlreiche Beispiele in Cavoukian: Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices
- Cavoukian, Popa: Privacy by ReDesign: A Practical Framework for Implementation
- Dennedy et al.: The Privacy Engineer's Manifesto Getting from Policy to Code to QA to Value (Chapters 3, 6)

c) Betreuer: Bastian Kemmler (LRZ)

d) Bearbeiter: noch offen

E. Special Use Cases

15 **Privacy and Car-to-Car Communication**

a) Inhalt:

Ad hoc Car-to-Car Communications enable the cooperation of vehicles by linking individual information distributed among multiple vehicles. The so-formed Vehicular Adhoc Network (VANET) works like a new „sensor“ increasing the drivers' range of awareness to spots which both the driver and onboard sensor systems otherwise cannot see. The C2C system electronically extends the driver's horizon and enables entirely new safety functions. While C2C communication systems enable cars to exchange data, anonymity of the vehicle and its driver must be protected to a level at least comparable to which users of mobile phones feel comfortable with today.

b) Literatur:

- Car2Car Communication Consortium: Manifesto: Overview of the C2C-CC System
- Dok et al.: Privacy Issues of Vehicular Ad-Hoc Networks
- de Fuentes et al.: Overview of security issues in Vehicular Ad-hoc Networks (Chapter 56 of Cruz-Cunha, Moreira (eds.): Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts)
- Rabieh et al.: A Secure and Privacy-Preserving Event Reporting Scheme for Vehicular Ad Hoc Networks

c) Betreuer: Matthias Maiterth
d) Bearbeiter: noch offen

---

16 **Privacy in Genomics**

a) Inhalt:
The protection of identity of participants in medical research has traditionally been guaranteed by the maintenance of the confidentiality of health information through mechanisms such as only releasing data in an aggregated form or after identifying variables have been removed. This protection of privacy is regarded as a fundamental principle of research ethics, through which the support of research participants and the public is maintained. Whilst this traditional model was adopted for genetics and genomics research, and was generally considered broadly fit for purpose, this approach seems to be increasingly untenable in genomics. Privacy risk assessments need to consider the whole data environment, not merely the quality of the dataset to be released in isolation. As sources of data proliferate, issues of privacy protection are increasingly problematic in relation to the release of genomic data.

b) Literatur:

- Heeney et al.: Assessing the Privacy Risks of Data Sharing in Genomics
- Humbert et al.: Interdependent privacy games: the case of genomics
- Telenti et al.: On genomics, kin, and privacy
- Erlich, Narayanan: Routes for breaching and protecting genetic privacy
- Gymrek et al.: Identifying Personal Genomes by Surname Inference
- Ayday et al.: Privacy-Preserving Processing of Raw Genomic Data

c) Betreuer: Karl Fürlinger
d) Bearbeiter: noch offen

---

17 **Privacy and Electronic Voting**

a) Inhalt:
Electronic voting (also known as e-voting) is voting using electronic systems to aid casting and counting votes. E-voting technology ranges from punched cards to specialized voting kiosks (including self-contained direct-recording electronic voting systems, or DRE). It can also involve transmission of ballots and votes via private computer networks, or the Internet. In general, e-voting includes supervised voting by representatives of governmental or independent electoral authorities and remote e-voting where voting is performed within the voter's sole influence and not physically supervised by representatives of governmental authorities. While e-voting technology can speed the counting of ballots, there has been contention that e-voting, especially DRE voting, could facilitate *electoral fraud*.

b) Literatur:

- Oo, Aung: A Survey of Different Electronic Voting Systems
- Zissis, Lekkas: Securing e-Government and e-Voting with an open cloud computing architecture
- Olembo, Volkamer: E-Voting System Usability: Lessons for Interface Design, User Studies, and Usability Criteria (Chapter 11 in Saeed, Reddick (eds.): Human-Centered System Design for Electronic Governance)

c) Betreuer: Karl Fürlinger
d) Bearbeiter: noch offen

---

18 **Privacy in Smart Grids**

a) Inhalt:
New technologies for computerized metering and data collection in the electrical power grid promise to create a more efficient, cost-effective, and adaptable smart grid. However, naive implementations of smart grid data collection could jeopardize the privacy of consumers, and concerns about privacy are a significant obstacle to the rollout of smart grid technology.

b) Literatur:

- Rial, Danezis: Privacy-Preserving Smart Metering
- Wang, Lu: Cyber security in the Smart Grid: Survey and challenges
- Petrlic: A privacy-preserving Concept for Smart Grids
- Mármol et al.: Privacy-enhanced architecture for smart metering
- Cavoukian: Smart Meters in Europe: Privacy by Design at its Best

c) Betreuer: Vitalian Danciu
d) Bearbeiter: noch offen

---

19 **Privacy and Workflow Provenance**

a) Inhalt:
Scientific workflow systems increasingly store provenance information about the module executions. However, authors/owners of workflows may wish to keep some of this information confidential. The problem to solve is the „secure-view“ problem.

b) Literatur:

- Davidson et al.: Preserving module privacy in workflow provenance
- Davidson et al.: Privacy Issues in Scientific Workflow Provenance
- Davidson et al.: Enabling Privacy in Provenance Aware Workflow Systems

c) Betreuer: Michael Schiffers

d) Bearbeiter: noch offen

---

20 **Privacy in Social Networks**

a) Inhalt:

With the proliferation of online social networks, there has been increasing concern about the privacy of individuals participating in them. While disclosing information on the web is a voluntary activity on the part of the users, users are often unaware of who is able to access their data and how their data can potentially be used. Data privacy is defined as „freedom from unauthorized intrusion“. However, what constitutes an unauthorized intrusion in social networks is an open question which is further complicated by two confusing facts: Firstly, our online communication is usually accessible to a vast number of people. In addition to these online friends as a „known audience“, there are other „unknown audiences“, such as advertisers who purchase the users' aggregated profile information from social media companies to address their target audiences. Secondly, many users appear not to feel threatened in terms of their need for and experiences of privacy when communicating online. Although they are aware of their data's publicity on an abstract level, many feel free to speak and to open up to others.

b) Literatur:

- Madejski et al.: A Study of Privacy Settings Errors in an Online Social Network
- Zheleva, Geetor: Privacy in Social Networks: A Survey (Chapter 10 in C. C. Aggarwal (ed.): Social Network Data Analytics)
- Beye et al.: Privacy in Online Social Networks
- Several chapters in Trepte, Reinecke (eds.): Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web

c) Betreuer: Nils gentschen Felde

d) Bearbeiter: noch offen

III. Zusammenfassung

In der folgenden Tabelle sind noch einmal kurz die Themen gelistet. Auf der jeweils angegebenen Seitenzahl finden Sie neben einer Mini-Inhaltsangabe auch Literaturhinweise und den Betreuer des Themas. Kontaktdaten können Sie den jeweiligen Links entnehmen.

Es wird *dringend* empfohlen, frühzeitig Kontakt mit den Betreuern aufzunehmen.