General Overview and Legal Frameworks

Introduction to Privacy Engineering

Daniel Aranki
University of California, Berkeley

General Overview and Legal Frameworks

Administrative: Objectives of the Course

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

What is the Course About?

- Engineering-applicable techniques to protect privacy in data
- Different settings and scenarios
- Privacy trade-offs
- Preserving utility from data

Objectives

By the end of the semester you will be able to:

- Describe the different technical paradigms of privacy applicable in engineering
- Critique the strengths and weaknesses of the different paradigms
- Implement the different privacy paradigms
- Keep up with the state-of-the art

Weekly Cycle

- Readings
- Asynchronous elements (use handouts)
- Self work
- Live session

General Overview and Legal Frameworks

A Brief History of Privacy: From Ancient Greece To Modern Photography and the Printing Press

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

Old Concepts of Privacy

- Aristotle's two spheres:
  - Public sphere (polis) – political life
  - Private sphere (oikos) – domestic life
- Attorney-client privilege
- Doctor-patient privilege

Concept(s) of Privacy

- Ability to seclude oneself
- Ability to express oneself selectively
- Physical privacy: one's space or solitude
  'The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures" – Fourth amendment

Other Aspects

- Trade-offs
  - "Those who would give up essential liberty to purchase a little temporary safety, deserve neither liberty nor safety." – Benjamin Franklin
  - Utility
  - Cost
  - Freedom of information
- Cultural context
- Time evolution
- Secrecy versus privacy

General Overview and Legal Frameworks

A Brief History of Privacy: Modern Photography and Beyond

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

The Right to Privacy

According to Warren and Brandeis (1890)[1]:

- "Right to life" evolved; expanded remedies: physical vs "sensation;" Examples: battery vs assault, slander and libel, intellectual property
- Domestic sphere (oikos) is being invaded by instantaneous photography and wide-spread press
- Remedies for circulating portraits of people? "Gossip" by newspapers?
- "The right to be left alone." – Judge Cooley
- New nuances of invasion of privacy

[1] Samuel D Warren and Louis D Brandeis (1890). "The right to privacy". In: **Harvard law review**, pp. 193–220.

The Integration of Information Systems

- Information systems are emerging (data banks)
- Lack of memory loss
- Four states of privacy[2]:
  1. Solitude: physical
  2. Intimacy: close relationship
  3. Anonymity: "public privacy"
  4. Reserve: psychological
- Even more nuance

[2] Alan F Westin (1968). "Privacy and freedom". In: **Washington and Lee Law Review** 25.1, p. 166.

General Overview and Legal Frameworks

A Brief History of Privacy: Artificial Intelligence and the Inference Threat

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

Inference Threat

Yet another wave of nuances:

- Information can be **inferred** about us; AI, statistical learning, etc...

Netflix Prize De-Anonymization

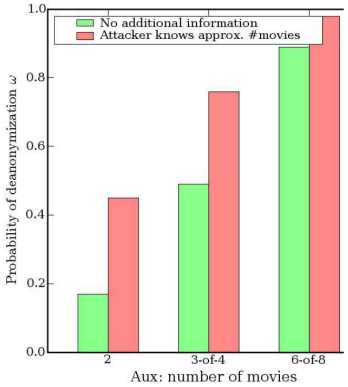

Figure: De-anonymization probability (Narayanan and Shmatikov, 2008)

Language identification in VoIP

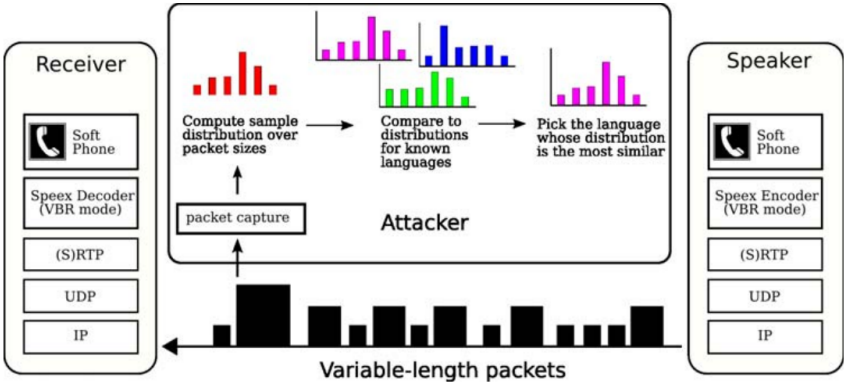

Figure: Attack setting (Wright et al., 2007)

Language identification in VoIP

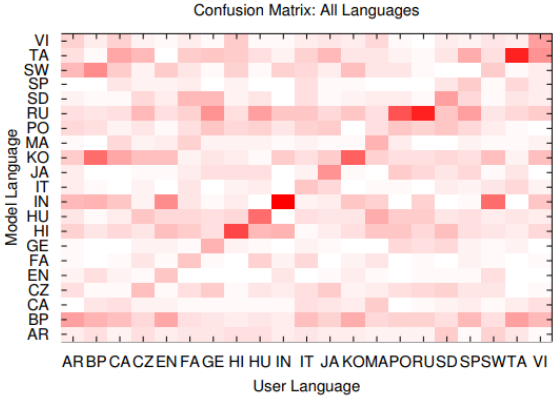

Figure: Attack results (Wright et al., 2007)

HTTPS: which page have you visited?

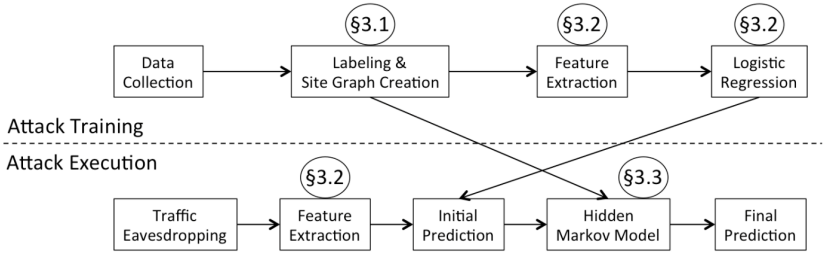

Figure: Attack pipeline (Miller et al., 2014)

HTTPS: Which Page Have You Visited?

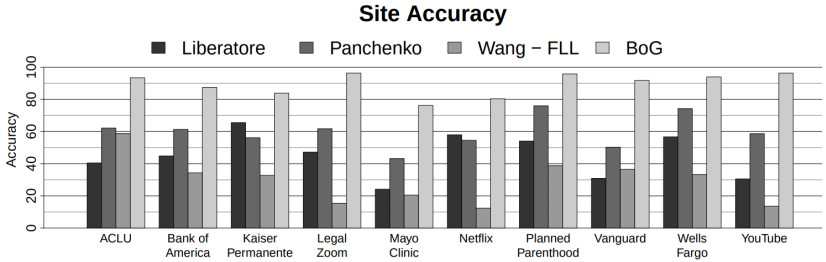

Figure: Attack results (Miller et al., 2014)

Utility vs Privacy

- Rule of thumb: as privacy protection grows, utility decreases
- Impossibility result in statistical databases[3]
- Perhaps achievable in other scenarios?

[3] Cynthia Dwork (July 2006). "Differential Privacy". In: **33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)**. Vol. 4052. Venice, Italy: Springer Verlag, pp. 1–12. ISBN: 3-540-35907-9. URL: https://www.microsoft.com/en-us/research/publication/differential-privacy/.

General Overview and Legal Frameworks

Legal Frameworks of Privacy: US Privacy Act of 1974

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

Motivation

- Watergate scandal: curb illegal surveillance and investigation
- Increasing use of databanks and computer systems

Features of the US Privacy Act of 1974

- Covers data about individuals (US citizens or "aliens lawfully admitted for permanent residence")[4]
- Applicable only to government agencies
- Commercial arena? Federal Trade Commission's Fair Information Practices
- "The right to privacy is a personal and fundamental right protected by the Constitution of the United States."
- Served as a model for privacy legislation worldwide

[4] Privacy Act (1974). "US Congress". In: **5 U.S.C.** §552a.

US Privacy Act Fair Information Practices

1. Openness and transparency
2. Individual participation
3. Collection limitation
4. Data quality
5. Use limitation
6. Reasonable security
7. Accountability

FTC's Fair Information Practices

- Federal Trade Commission report on online privacy[5]
- Notice/Awareness
- Choice/Consent
- Access/Participation
- Integrity/Security
- Enforcement/Redress

[5] Robert Pitofsky et al. (1998). **Privacy online: A report to congress**. Commission Findings. Federal Trade Commission.

General Overview and Legal Frameworks

Legal Frameworks of Privacy: General Data Protection Regulation (GDPR)

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

Driving Force and Scope

- Driving concept: Privacy is a fundamental human right.
- Primary motivation: Adapt to changes in the data ecosystem
- Covers personal data of all people residing in the EU by any data collector or processor

Features of the GDPR

- Opt-in and consent
- Right to access
- Right to be forgotten
- Liability includes processors as well as controllers
- Data Protection Officer
- Regulation of design and retention
- Security: impact assessment
- Breach procedures: notification and penalties
- Transparency
- Age protection: minimum age is 16

Comparing GDPR to the US Privacy Act of 1974

- Scope: GDPR is broader, covering all individuals in the EU, applicable to industry and government (controllers and processors).
- GDPR consent concept is stronger.
- GDPR provides the right to be forgotten.
- GDPR requires documentation and a designated "Data Protection Officer."

General Overview and Legal Frameworks

Course Overview: What Will This Course Cover

Introduction to Privacy Engineering
Daniel Aranki
University of California, Berkeley

Publishing Types

Microdata: Detailed records, each of an entity (person, company, etc.)

Macrodata: Derived statistics from the dataset

Interactive: Can be queried

Noninteractive: A snapshot is released

Utility Landscape

- Databases
- Data mining
- Information disclosure
- Learning and inference

Privacy Threats Landscape

Membership Disclosure: being able to tell that a person is in (or not in) a dataset (confidentiality)

Identity Disclosure: being able to tell the identity of the person to whom the record corresponds (anonymity)

Inference Threat: being able to tell that a person has a specific (sensitive) attribute:

- Attribute disclosure.
- Inference of undisclosed attributes.

Membership Disclosure

- Being able to tell that a person is in (or not in) a dataset (confidentiality)

Gender	Age	Group?
Male	[31-35]	Treatment
Male	[31-35]	Control
Male	[31-35]	Control
Male	[31-35]	Treatment
Female	[26-30]	Control
Female	[26-30]	Control
Female	[26-30]	Treatment
Female	[26-30]	Treatment

Identity Disclosure

- Being able to tell the identity of the person to whom the record corresponds (anonymity)

Gender	**Age**	**Medical Condition**	**Fully paid bill?**
Male	[31-35]	Back injury	Yes
Male	[36-40]	Flu	No
Male	[31-35]	Cancer	Yes
Male	[31-35]	Healthy	No
Female	[26-30]	Flu	No
Female	[26-30]	Sprained ankle	No
Female	[26-30]	Back injury	Yes
Female	[26-30]	Sprained ankle	Yes

Attribute Disclosure

- Being able to tell that a person has a specific (sensitive) attribute

Gender	Age	Medical Condition	Fully paid bill?
Male	[31-35]	Flu	Yes
Male	[31-35]	Flu	No
Male	[31-35]	Flu	Yes
Male	[31-35]	Flu	No
Female	[26-30]	Flu	No
Female	[26-30]	Sprained ankle	No
Female	[26-30]	Back injury	Yes
Female	[26-30]	Sprained ankle	Yes

Inference Threat

- Being able to tell something new (undisclosed) about a person

Movie	**Like/Dislike**
Fahrenheit 9/11	Like
Inside Job	Like
Fahrenhype 9/11	Dislike
2016: Obama's America	Dislike

Course Overview

- Privacy by Design
- Background knowledge: probability theory, information theory and machine learning
- Randomized Response (Warner, 1965)
- k-Anonymity (Sweeney, 2002)
- ℓ-Diversity (Machanavajjhala et al., 2007)
- t-Closeness (N. Li, T. Li, and Venkatasubramanian, 2007)
- δ-Presence (Nergiz, Atzori, and Clifton, 2007)
- ϵ-Differential Privacy (Dwork, 2006)
- Honest but curious (Pin Calmon and Fawaz, 2012)
- Private Disclosure of Information (Aranki and Bajcsy, 2015)

References I

- Aranki, Daniel and Ruzena Bajcsy (2015). "Private Disclosure of Information in Health Tele-monitoring". In: *arXiv preprint arXiv:1504.07313*.
- Dwork, Cynthia (July 2006). "Differential Privacy". In: *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*. Vol. 4052. Venice, Italy: Springer Verlag, pp. 1–12. ISBN: 3-540-35907-9. URL: https://www.microsoft.com/en-us/research/publication/differential-privacy/.
- Li, N., T. Li, and S. Venkatasubramanian (Apr. 2007). "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity". In: *2007 IEEE 23rd International Conference on Data Engineering*, pp. 106–115. DOI: 10.1109/ICDE.2007.367856.
- Machanavajjhala, Ashwin et al. (Mar. 2007). "L-diversity: Privacy Beyond K-anonymity". In: *ACM Trans. Knowl. Discov. Data* 1.1, pp. 3–54. ISSN: 1556-4681. DOI: 10.1145/1217299.1217302. URL: http://doi.acm.org/10.1145/1217299.1217302.
- Miller, Brad et al. (2014). "I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis". In: *Privacy Enhancing Technologies*. Ed. by Emiliano De Cristofaro and Steven J. Murdoch. Cham: Springer International Publishing, pp. 143–163. ISBN: 978-3-319-08506-7.
- Narayanan, Arvind and Vitaly Shmatikov (2008). "Robust de-anonymization of large sparse datasets". In: *Security and Privacy, 2008. SP 2008. IEEE Symposium on*. IEEE, pp. 111–125.

References II

- Nergiz, Mehmet Ercan, Maurizio Atzori, and Chris Clifton (2007). "Hiding the Presence of Individuals from Shared Databases". In: *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*. SIGMOD '07. Beijing, China: ACM, pp. 665–676. ISBN: 978-1-59593-686-8. DOI: 10.1145/1247480.1247554. URL: http://doi.acm.org/10.1145/1247480.1247554.
- Pin Calmon, F. du and N. Fawaz (Oct. 2012). "Privacy against statistical inference". In: *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1401–1408. DOI: 10.1109/Allerton.2012.6483382.
- Pitofsky, Robert et al. (1998). *Privacy online: A report to congress*. Commission Findings. Federal Trade Commission.
- Privacy Act (1974). "US Congress". In: *5 U.S.C.* §552a.
- Sweeney, Latanya (2002). "k-anonymity: A model for protecting privacy". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05, pp. 557–570.
- Warner, Stanley L (1965). "Randomized response: A survey technique for eliminating evasive answer bias". In: *Journal of the American Statistical Association* 60.309, pp. 63–69.
- Warren, Samuel D and Louis D Brandeis (1890). "The right to privacy". In: *Harvard law review*, pp. 193–220.
- Westin, Alan F (1968). "Privacy and freedom". In: *Washington and Lee Law Review* 25.1, p. 166.

References III


Wright, Charles V. et al. (2007). "Language Identification of Encrypted VoIP Traffic: Alejandra Y Roberto or Alice and Bob?" In: *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*. SS'07. Boston, MA: USENIX Association, 4:1–4:12. ISBN: 111-333-5555-77-9. URL: http://dl.acm.org/citation.cfm?id=1362903.1362907.