Design Principles for Privacy in the Internet of Things

Roy Dong

University of California, Berkeley

Shaunak D. Bopardikar

United Technologies Research Center

Helsinki Privacy Experiment

- 10 households (12 individuals) monitored over 6 months.
- 3-5 video cameras with microphones, computer keylogging and screenshots, wireless and wired network, smartphone, TV and DVD, customer loyalty cards.

HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY

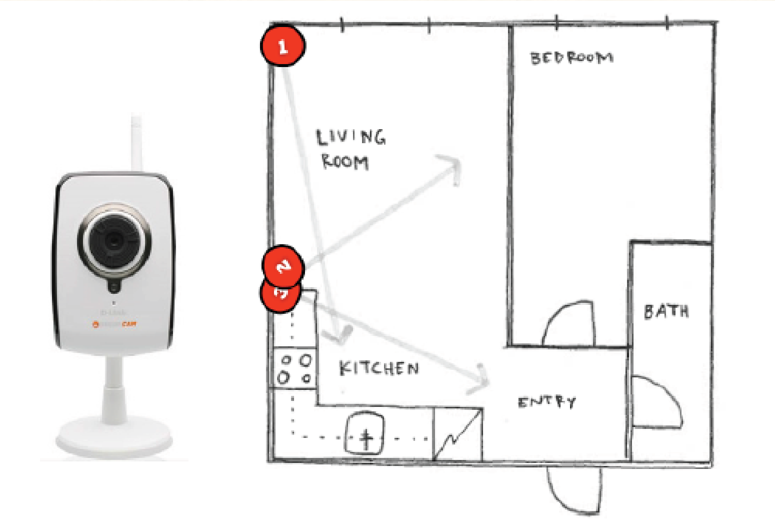


Figure 2. The Wi-Fi camera and an example positioning. The plan was redrawn from a subject's original.

[Long-term Effects of Ubiquitous Surveillance in the Home (2012)]

Helsinki Privacy Experiment

- **Results**:
  - Habituation
  - All but 1 participant showed privacy-seeking behavior: ceasing a behavior entirely, hiding things, acting privately, manipulating sensors. Known as the **chilling effect**.

[Long-term Effects of Ubiquitous Surveillance in the Home (2012)]

Outline

- Privacy
  - What’s at stake?
- Privacy by Design
  - **Passive** privacy analysis
  - **Active** privacy mechanisms
  - **Optimal** privacy design
- Industrial Need for Privacy-Preserving Mechanisms

Privacy by Design

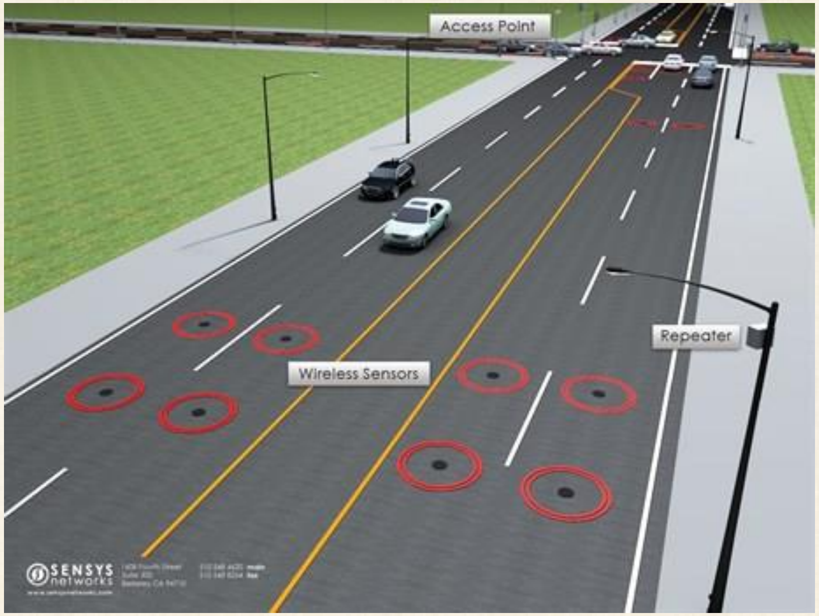
- **Passive** privacy analysis
  - For a fixed system, **quantify** the privacy risk of users.

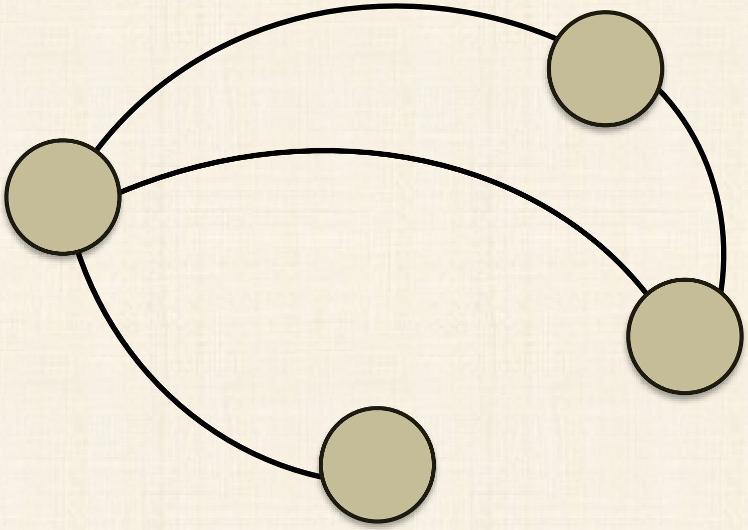
Privacy by Design

- **Passive** privacy analysis

Example:

- **RD**, Krichene, Bayen, Sastry, “Differential Privacy of Populations in Routing Games” (2015)
  - Given traffic infrastructure, learning dynamics, and a noise model, calculate the level of differential privacy.



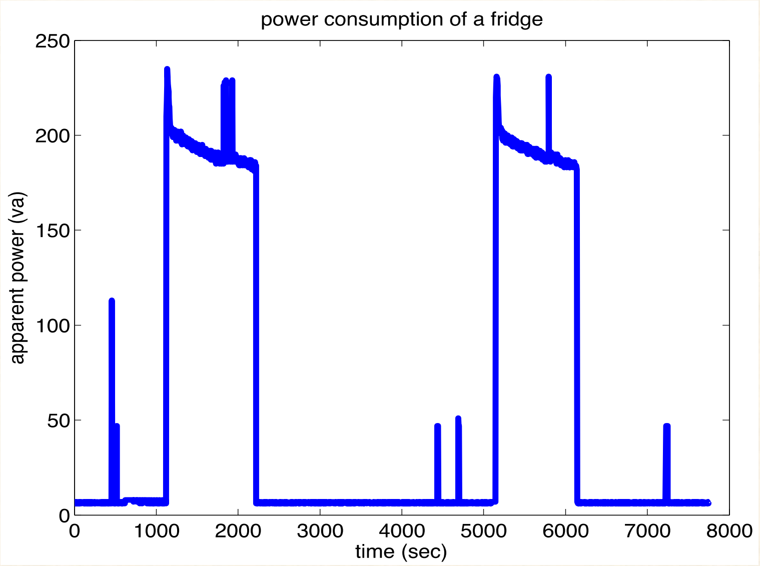


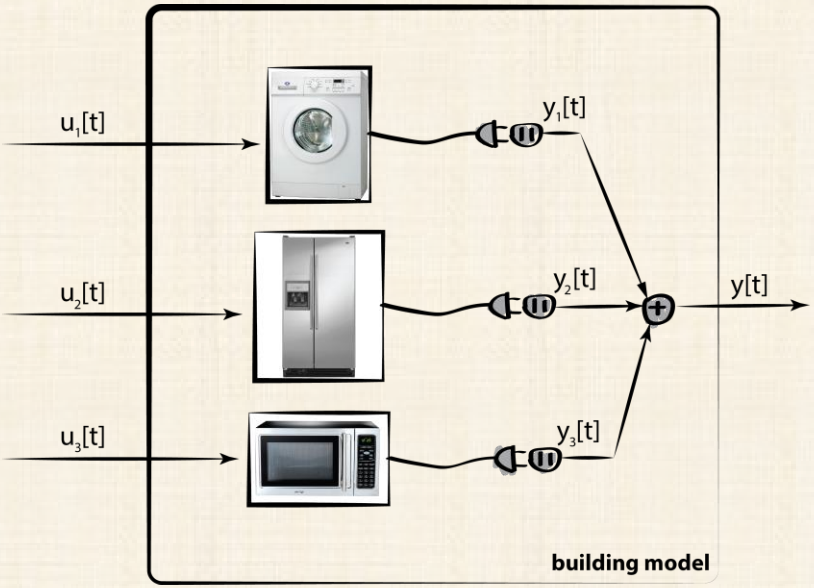
Privacy by Design

- **Passive** privacy analysis

Example:

- **RD**, Ratliff, Ohlsson, Sastry, “Fundamental Limits of Nonintrusive Load Monitoring” (2014)
  - Given device dynamics, quantify inherent uncertainty in energy disaggregation problem.





Privacy by Design

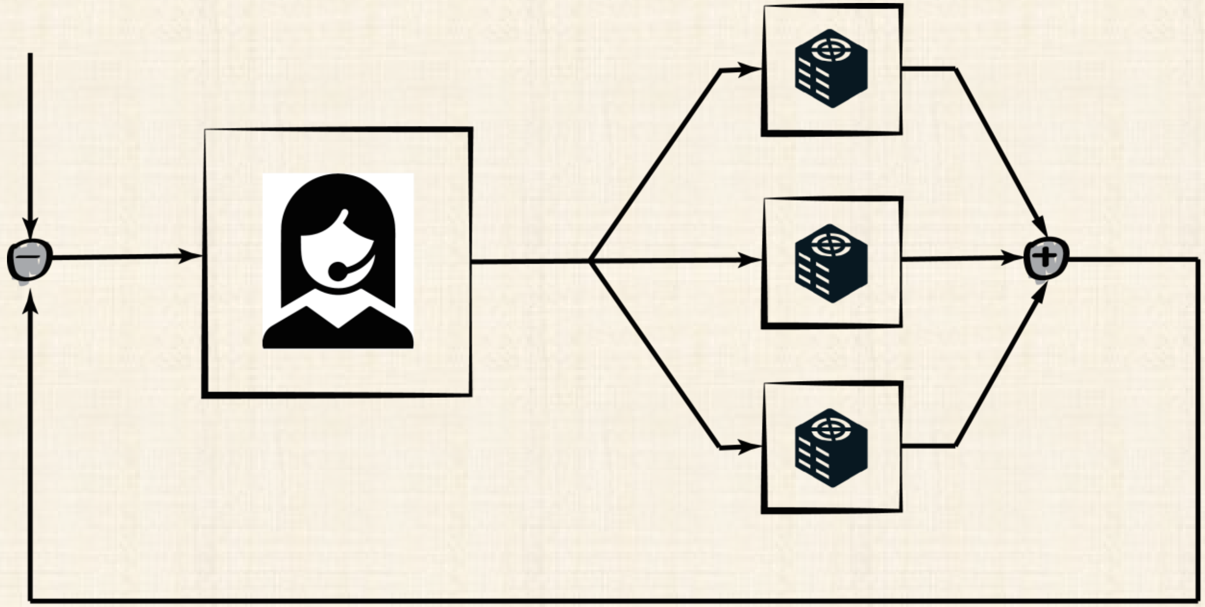
- **Active** privacy mechanisms
  - Fix a **parameterized** privacy-preserving scheme.
  - Pick the privacy parameter to best **trade-off** the utility of the collected data with the privacy of users.

Privacy by Design

- **Active** privacy mechanisms

Example:

- **RD**, Cárdenas, Ratliff, Ohlsson, Sastry, “Quantifying the Utility-Privacy Tradeoff in the Internet of Things,” (under review)
  - Pick a sampling frequency to tradeoff direct load control performance and user privacy.



Privacy by Design

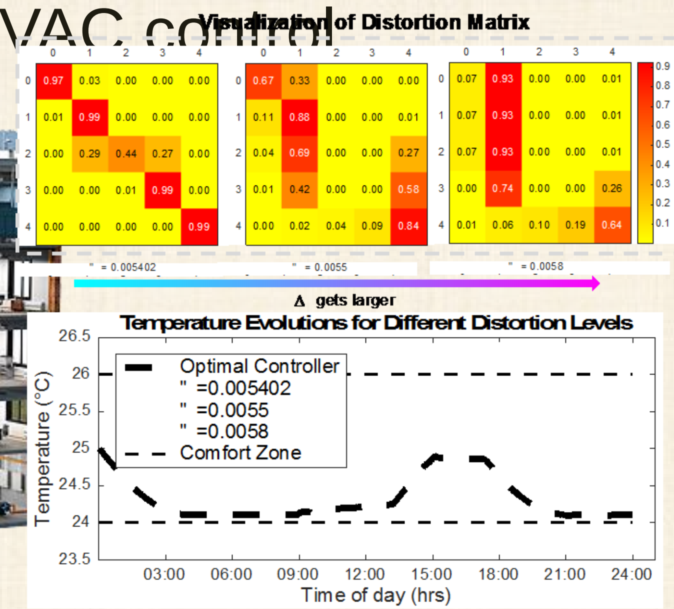
- **Optimal** privacy design
  - Fix **performance** metrics and **privacy** metrics.
  - Design a privacy-preserving mechanism that **maximizes** privacy, subject to performance **constraints**.

Privacy by Design

- **Optimal** privacy design

Example:

- Jia, **RD**, Sastry, Spanos, , Ratliff, Ohlsson, Sastry, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control,” (under review)
  - Minimize mutual information between individual traces and reported data, while still providing improved occupancy-based HVAC control



Privacy by Design

- **Passive** privacy analysis
- **Active** privacy mechanisms
- **Optimal** privacy design

Privacy-Awareness in Applications

Companies collect **data** from customers to recommend **maintenance schedules**

Pratt and Whitney

UTC Enterprise

Lenel (Access Control)

- ***Aerospace:***

From Customer Data to:
  - ***Mission History?***
  - ***Operator Usage?***

- ***Commercial:***

From Customer Data to:
  - ***User preferences?***
  - ***Occupancy patterns?***

- Multiple customers sharing their data (mix of public and private/proprietary)
- Access to “private” data would often lead to improved analytics
- Insight into customer perspective toward privacy

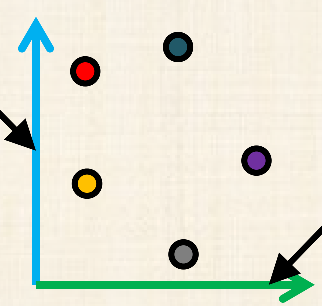
Other related examples:
- Automotive and Auto-insurance companies (Ref: NY times, Aug 15, 2014)
- Authentication based on gait (DHS CASTRA project, PI: Dr. Manikantan Shila, UTRC)

UTRC's Algebraic Topological Perspective to Privacy

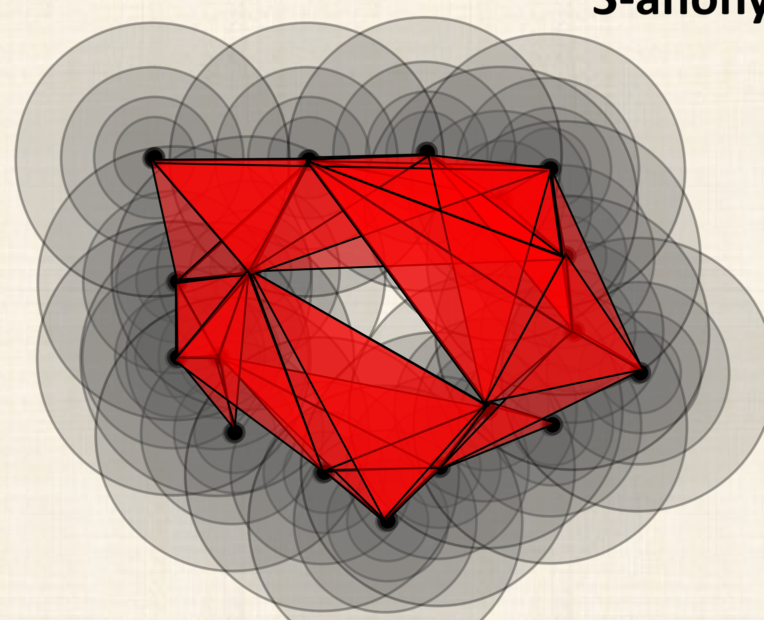
Age	ZIP Code	Salary
25	47677	$47,000
22	47602	$32,000
24	47678	$52,000
43	47905	$151,000
52	47909	$145,000
38	47906	$98,000
47	47605	$110,000
36	47673	$92,000
32	47607	$115,000

Age	ZIP Code
[22-25]	[47602-47678]
[22-25]	[47602-47678]
[22-25]	[47602-47678]
[38-52]	[47905-47909]
[38-52]	[47905-47909]
[38-52]	[47905-47909]
[32-47]	[47605-47603]
[32-47]	[47605-47603]
[32-47]	[47605-47603]

Data "at rest"

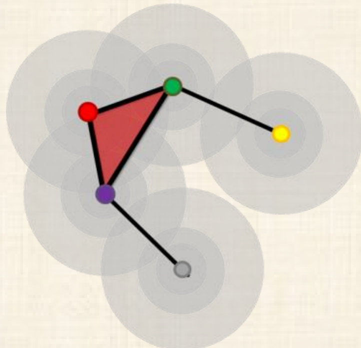


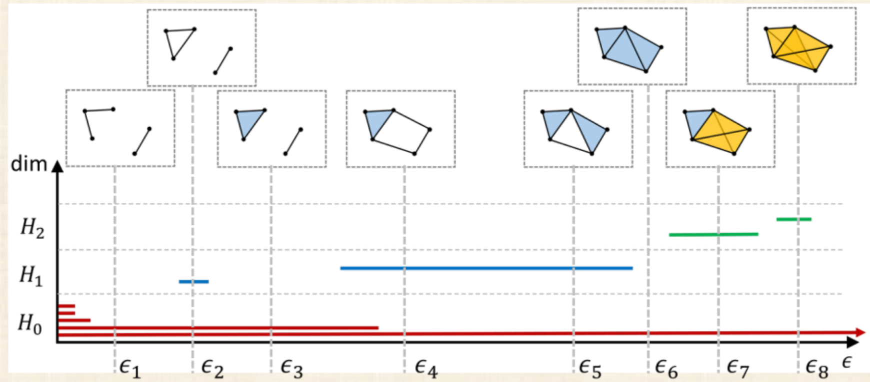
3-anonymity



*A. Speranzon and **S. D. Bopardikar**, "An Algebraic Topological perspective to Privacy", American Control Conference, 2016.*

UTRC's Algebraic Topological Perspective to Privacy

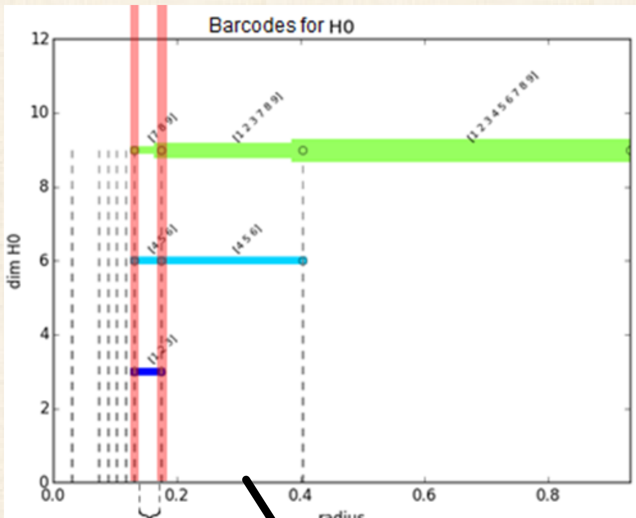




Bar code diagram

Sample results

Age	ZIP Code	Salary
25	47677	$47,000
22	47602	$32,000
24	47678	$52,000
43	47905	$151,000
52	47909	$145,000
38	47906	$98,000
47	47605	$110,000
36	47673	$92,000
32	47607	$115,000



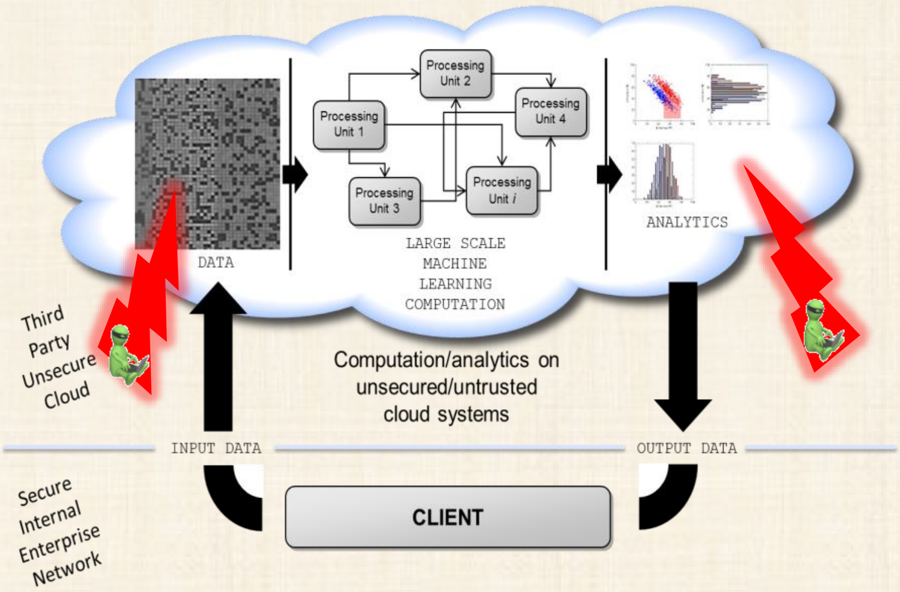
Age	ZIP Code	Salary
25	47677	$47,000
22	47602	$32,000
24	47678	$52,000
43	47905	$151,000
52	47909	$145,000
38	47906	$98,000
47	47605	$110,000
36	47673	$92,000
32	47607	$115,000

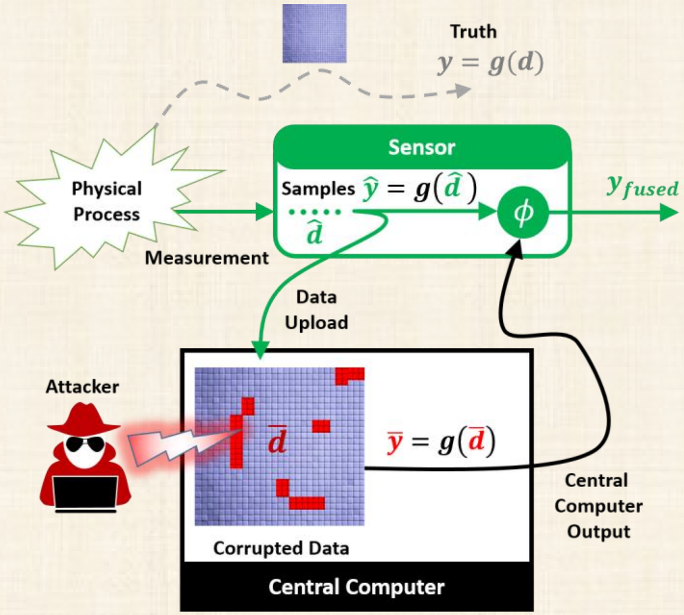
3-anonymity with most # of classes

***Extensions**: Categorical data, mixed continuous and categorical data, etc.*

*A. Speranzon and **S. D. Bopardikar**, "An Algebraic Topological perspective to Privacy", American Control Conference, 2016.*

Trusted Computation





- **Our approach**: Problem from Trusted Computation + Mathematics from *Adversarial Machine Learning*
- *Game-theoretic (iterative) methods* to produce a fusion solution that requires low complexity
- Theoretical conditions on convergence [Bopardikar et al, ACC 2015 and Automatica 2017]

- **Open directions:** joint privacy of data and security of computation, distributed repetitive games

Prototypical (Abstract) problem

- Compute $y = F(x, p)$
  - x: public variables
  - p: private variables (or functions)
  - F: algorithm/code which could be partly private
    - Subroutines could be proprietary
  - y: useful output for a legitimate/honest user
- Goal: prevent reverse engineering of p, F
- Features:
  - Accuracy is ***very important***!
  - Protection against ***multiple runs*** of the code
  - Probabilities are ***not provided*** as specifications!

Conclusion – Takeaways, Gaps

- Privacy problems often solved through contracts
  - Binary (opt in/out)
  - Protect confidentiality

- Privacy metrics need to be more visual/psychological
  - Very little intuition behind value of ϵ in differential privacy
  - How do we verify privacy guarantees?

- Privacy interlinks/conflicts with security in many scenarios
  - Cyber tools are necessary, but not sufficient
  - Security problem can be difficult under privacy constraints

- Current trends toward video-streams
  - Computer vision, data analytics, dynamical systems