Privacy-Aware Distributed Detection

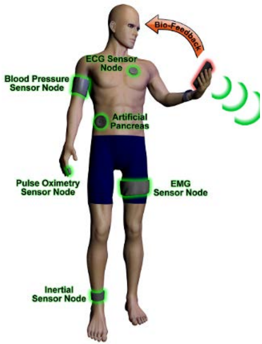
Tobias J. Oechtering

joint work with Zuxing Li

KTH Royal Institute of Technology,
School of EE and ACCESS Linnaeus Center,
Communication Theory Lab,
Stockholm, Sweden

Princeton, June 27, 2014

Physical-Layer Privacy for E-Health



- Distributed detection for health monitoring - two concerns:
  - Detection performance
  - Privacy risk
- **Privacy-per-design approach:** Include both concerns in the system design!
  - Privacy-aware distributed detection

Benefits: Enhancement of existing privacy schemes, and/or ensuring privacy when *existing schemes cannot be applied*, e.g. statistical inference attack

- Interesting for many other IoT/cyber-physical applications.

Related Literature

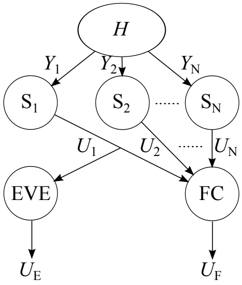
- **Distributed detection.** Well established theory, many substantial contributions in the 80's and 90's.
  - [Tenney, Sandell Jr.,'81] introduced Bayesian problem
- **Physical-layer security.** A hot topic in the last decade.
  - [Shannon,'49] introduced communication theory of secrecy systems.
- Recently, **physical-layer security in distributed detection.**
  - Perfect secrecy using *KL divergence* as security metric in the *asymptotic regime* in the number of sensors:
    - [Marano et al.,'09][1] Eavesdrooper (Eve) intercepts wireless transmissions from remote sensors to infer on natures state as well
    - [Nadendla et al.,'10][2] Eve intercepts sensors digital data
  - Others deal with Byzantine attacks in distributed detection

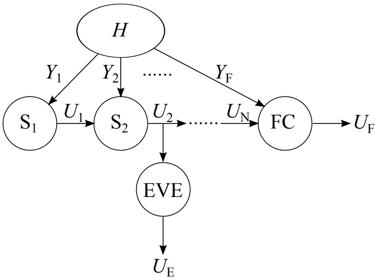
[1] S. Marano, V. Matta, and P. K. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *IEEE Trans. Signal Processing*, vol. 57, no. 5, pp. 1976-1986, 2009.

[2] V. S. S. Nadendla, H. Chen, and P. K. Varshney, "Secure distributed detection in the presence of eavesdroppers," in *Proc. of ASILOMAR 2010*, 2010, pp. 1437-1441.

Distributed Detection Vulnerable to an Eavesdropper

We keep N fixed and Eve wants to detect H as well!





- **Binary** hypothesis H and decisions U_k
- **Conditionally independent** observations Y_k given H
- The eavesdropper is known to intercept a local decision.

Parallel Distributed Detection with an Eavesdropper

Independently randomized decision strategies at

- remote sensors

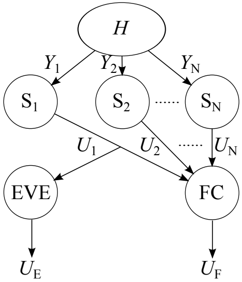
$$\gamma_i(y_i) = U_i$$

- fusion center

$$\gamma_{\mathsf{F}}(u_1, \dots, u_{\mathsf{N}}) = U_{\mathsf{F}},$$

- eavesdropper

$$\gamma_{\mathsf{E}}(u_1) = U_{\mathsf{E}}.$$



Eavesdropper is **informed** about the system and **greedy**.

Bayesian vs. Neyman-Pearson Approach

- **Bayesian approach**: Minimize the Bayesian risk
  - Known prior probability $p_H(h)$
  - Assign detection costs $c_{U_\mathsf{F},H}(u_\mathsf{F},h)$.
  - Bayesian risk of the fusion node $c_\mathsf{F} = \sum_{u_\mathsf{F},h} p_{U_\mathsf{F},H}(u_\mathsf{F},h) c_{U_\mathsf{F},H}(u_\mathsf{F},h)$
- **Neyman-Pearson approach**: Maximize detection probability $p_\mathsf{F}^\mathsf{D} = p_{U_\mathsf{F}|H}(1|1)$ with an upper bound on the false alarm probability $p_\mathsf{F}^\mathsf{F} = p_{U_\mathsf{F}|H}(1|0)$

Questions: How to extend problems to include an eavesdropper? What are (properties of) optimal decision strategies? ...

- [ICC'14][3] Privacy-constrained parallel Bayesian setting.
- [ICC'14 workshop][4] Corresponding Neyman-Pearson setting.

[3] Z. Li, T. J. Oechtering, and K. Kittichokechai, "Parallel distributed Bayesian detection with privacy constraints," in *Proc. IEEE ICC 2014*.

[4] Z. Li, T. J. Oechtering, and J. Jaldén, "Parallel distributed Neyman-Pearson detection with privacy constraints," in *Proc. IEEE ICC 2014 Workshop*.

Privacy-Constrained Bayesian Detection Problem

- Bayesian approach:
  - Define costs for Eve $c_{U_\mathsf{E},H}(u_\mathsf{E},h)$
  - Assume Eve knows prior probability $p_H(h)$
- Privacy metric (**minimal** Bayesian risk, since Eve is greedy):

$$c_\mathsf{E}^{\min} = \min_{\gamma_\mathsf{E}} c_\mathsf{E} = \min_{\gamma_\mathsf{E}} \sum_{u_\mathsf{E},h} p_{U_\mathsf{E},H}(u_\mathsf{E},h) c_{U_\mathsf{E},H}(u_\mathsf{E},h).$$

  - A detection-theoretic operational privacy metric!

Privacy-constrained parallel distributed Bayesian detection problem

$$\min_{\gamma_1,\gamma_2,\cdots,\gamma_\mathsf{N},\gamma_\mathsf{F}} c_\mathsf{F}, \quad \text{s.t.} \quad c_\mathsf{E}^{\min} \geq \beta.$$

Person-by-Person Optimality

- Properties of local person-by-person optimal decision tests are **necessary** to be satisfied by the global optimal tests.

Privacy-constrained person-by-person optimization of γ_1

$$\min_{\gamma_1} c_{\mathsf{F}}, \quad \text{s.t.} \quad c_{\mathsf{E}}^{\min} \geq \beta,$$

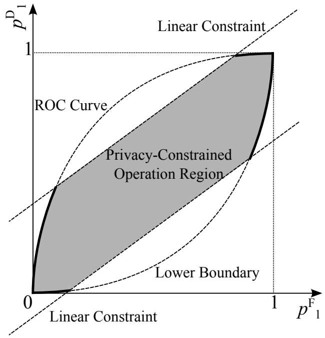
while all other decision strategies are fixed.

Observations:

- Strategy γ_1 determines operation point $(p_1^{\mathsf{F}}, p_1^{\mathsf{D}})$.
  - **Objective** $c_{\mathsf{F}}(p_1^{\mathsf{F}}, p_1^{\mathsf{D}}) = a_1 p_1^{\mathsf{F}} + b_1 p_1^{\mathsf{D}} + c_1$ is **linear** in $(p_1^{\mathsf{F}}, p_1^{\mathsf{D}})$.
  - **Constraints** $c_{\mathsf{E}}^{\min} \geq \beta \Leftrightarrow c_{\mathsf{E}}(p_1^{\mathsf{F}}, p_1^{\mathsf{D}}) \geq \beta, \forall \gamma_{\mathsf{E}}$ are **linear** in $(p_1^{\mathsf{F}}, p_1^{\mathsf{D}})$.

Illustration of Privacy-Constrained PBPO

- A linear objective over a convex set:



Person-by-Person Optimality

It is **sufficient** to consider operating points $(p_1^\mathsf{F}, p_1^\mathsf{D})$ on the **bold boundary sections** .

Deterministic LRT Optimality

- Since the curved boundary is achieved by likelihood ratio tests (LRTs) assuming observations Y_1 with continuous support:

Theorem

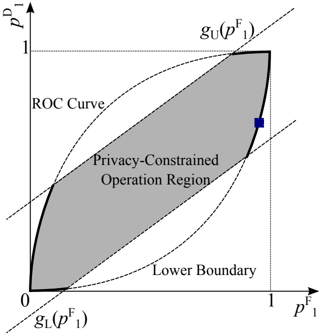
It is **sufficient** to consider **deterministic** likelihood ratio tests (LRTs) for the local person-by-person optimal and global optimal decision strategies of the eavesdropped decision maker (DM) S_1.

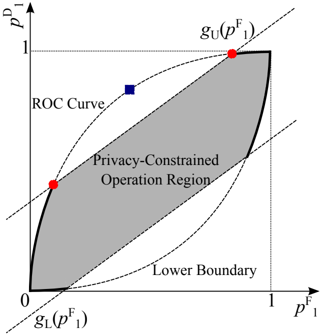
Remark:
Same holds for other decision strategies as well.

Extended Privacy-Constrained PBPO Algorithm

Remark

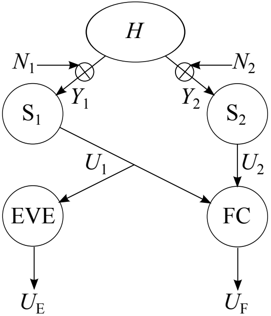
The algorithmic method of PBPO[6] can be easily extended to incorporate the privacy constraint.





[6] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 995-1000, 1989.

AWGN Example



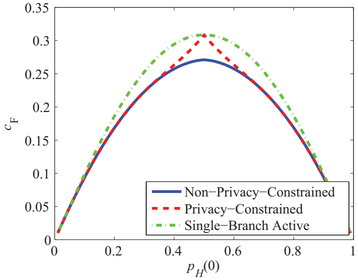
- Independent $N_i \sim \mathcal{N}(0,1)$
- Bayesian costs such that c_{F} and $c_{\mathrm{E}}^{\min}$ measure **average detection error** probabilities.

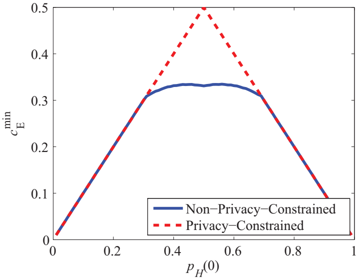
Maximal privacy constraint - Interception should not improve Eves risk compared to the risk based on prior knowledge only!

- Can be achieved by cutting of sensor with intercepted link!
- **Question:** Can we do better?

Tradeoff: Detection vs. Privacy Performance

- **Answer:** Yes! Intercepted local decision can be useless for Eve, but useful for fusion center due to information from other remote sensor!





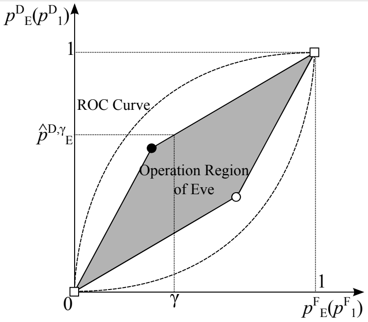
Privacy-Constrained Neyman-Pearson Problem

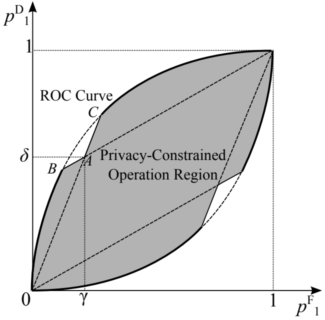
- Privacy metric (based on the Neyman-Pearson criterion):

$$\hat{p}_{\mathsf{E}}^{\mathsf{D},\gamma} = \max_{\gamma_{\mathsf{E}}} p_{\mathsf{E}}^{\mathsf{D}}, \quad \text{s.t.} \quad p_{\mathsf{E}}^{\mathsf{F}} \leq \gamma.$$

Privacy-constrained Neyman-Pearson problem

$$\max_{\gamma_1,\gamma_2,\gamma_{\mathsf{F}}} p_{\mathsf{F}}^{\mathsf{D}}, \quad \text{s.t.} \quad p_{\mathsf{F}}^{\mathsf{F}} \leq \lambda, \hat{p}_{\mathsf{E}}^{\mathsf{D},\gamma} \leq \delta.$$





Deterministic LRT Optimality for Remote DMs

Theorem

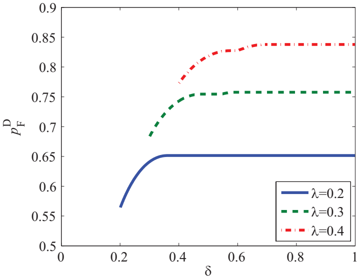
When a proper **randomized** fusion strategy is employed, it is **sufficient** to consider a **deterministic** LRT for each remote DM in the optimal privacy-constrained design.

For a design with **deterministic** strategies and $\gamma = \lambda$,

- $p_{\mathsf{F}}^{\mathsf{D}}$ increases along line segments $A \to B$ and $A \to C$ so that the optimal operating point is on the curved boundary, therefore
- it is **sufficient** to consider **deterministic** LRT for remote DMs.

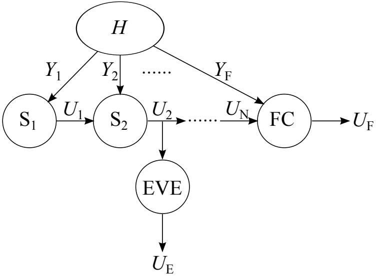
Detection - Privacy Tradeoff

- AWGN example, same settings as before



- The non-smooth curves result from using deterministic strategies at fusion node only.

Serial Setting with Privacy Constraint [ICASSP '14]



Similar to parallel setting:

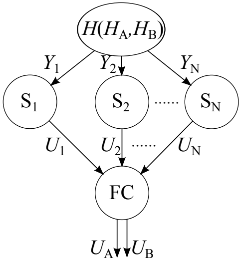
- same concepts and privacy metric
- similar problem formulation and conclusions

Major difference:

- Decision strategies $\gamma_i(y_i, u_{i-1}) = U_i$ are parametrized by previous decision u_{i-1} requires extension of analysis.

[7] Z. Li and T. J. Oechtering, "Tandem distributed Bayesian detection with privacy constraints," in *Proc. IEEE ICASSP 2014*, 2014, pp. 8188-8192.

Differential Privacy in Distributed Detection [Fusion'14]



- 4-ary hypothesis $H = (H_{\mathsf{A}}, H_{\mathsf{B}})$
  - **public binary** H_{A}
  - **private binary** H_{B}
- Fusion center
  - has access to all local decisions U_i,
  - should infer H_{A} while H_{B} should be kept private.

Parallel distributed Bayesian detection with a differential privacy constraint

$$\min_{\gamma_1,\gamma_2,\cdots,\gamma_{\mathsf{N}},\gamma_{\mathsf{A}}} c_{\mathsf{A}}, \quad \text{s.t.} \quad c_{\mathsf{B}}^{\min} \geq \beta.$$

[8] Z. Li and T. J. Oechtering, "Differential privacy in parallel distributed Bayesian detections," accepted at *Fusion 2014*, July 2014.

Optimality of Deterministic and Randomized LLCT(s)

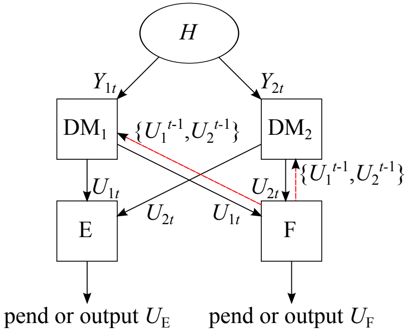
- Same conceptual tools are used as previously.
- Operation region is extended to 4-dimensions.
- More linear privacy constraints.

Theorem

It is **sufficient** to consider a **deterministic** linear likelihood combination test (LLCT) or a **randomized** strategy of LLCT. Randomized strategies are needed if operation point is determined by privacy constraints only.

- LLCT: $a_i f_{Y_i|H_A,H_B}(y_i|0,0) + b_i f_{Y_i|H_A,H_B}(y_i|1,0) + c_i f_{Y_i|H_A,H_B}(y_i|0,1) + d_i f_{Y_i|H_A,H_B}(y_i|1,1) \underset{u_i=0}{\overset{u_i=1}{\gtrless}} 0$

Sequential Detection with an Eavesdropper [GlobalSIP'14]



- Binary
  - hypothesis H and
  - decisions $U_{1,t}$, $U_{2,t}$, U_F, U_E
- $Y_1^T - H - Y_2^T$, each i.i.d. in time
- **Fusion** decides to **terminate sequential detection** system to make final decision U_F.
  - Finite-time horizon T

[10] Z. Li and T. J. Oechtering, "Privacy-concerned parallel distributed Bayesian sequential detection," invited to *IEEE GlobalSIP 2014*, December 2014.

Privacy-Concerned Detection Problem

- Independently randomized local decision strategies:

$$\gamma_{it}(y_{it}, u_1^{t-1}, u_2^{t-1}) = U_{it}$$

- $\gamma_{\mathsf{F}}, \gamma_{\mathsf{E}}$ are deterministic sequential detection strategies.

- **Privacy-concerned** Bayesian risk:

$$c_{\mathsf{P}} = \alpha c_{\mathsf{F}} - (1-\alpha) c_{\mathsf{E}}^{\min}, \ \alpha \in [0,1].$$

- Privacy-concerned parallel distributed Bayesian sequential detection problem:

$$\min_{\gamma_1^T, \gamma_2^T, \gamma_{\mathsf{F}}} c_{\mathsf{P}}.$$

Privacy-Concerned Person-by-Person Optimality

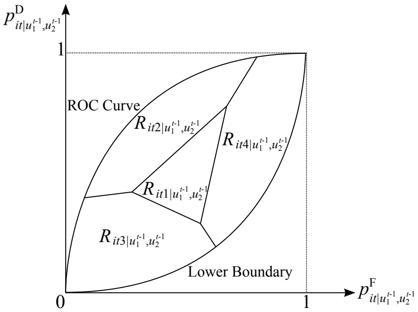
$$\min_{k\in\{1,\dots,K\}} \underbrace{\min_{(p^{\mathsf{F}}_{it|u_1^{t-1},u_2^{t-1}},p^{\mathsf{D}}_{it|u_1^{t-1},u_2^{t-1}})\in\mathcal{R}_{itk|u_1^{t-1},u_2^{t-1}}} \alpha c_{\mathsf{F}} - (1-\alpha) c_{\mathsf{E}k},}_{\textbf{convex optimization}}$$

with **convex set**

$$\mathcal{R}_{itk|u_1^{t-1},u_2^{t-1}} = \left\{ (p^{\mathsf{F}}_{it|u_1^{t-1},u_2^{t-1}}, p^{\mathsf{D}}_{it|u_1^{t-1},u_2^{t-1}}) \middle| \exists \gamma_{it|u_1^{t-1},u_2^{t-1}} \text{ with } \begin{array}{c} c_{\mathsf{E}k} \le c_{\mathsf{E}1} \\ \vdots \\ c_{\mathsf{E}k} \le c_{\mathsf{E}K} \end{array} \right\},$$

and c_{F}, $c_{\mathsf{E}k}$are **linear functions** of $p^{\mathsf{F}}_{it|u_1^{t-1},u_2^{t-1}}$, $p^{\mathsf{D}}_{it|u_1^{t-1},u_2^{t-1}}$.

Illustration of Sub-Regions with 4 Candidates of γ_{E}^{*}



- When privacy-concerned person-by-person optimizing γ_{F}, use the dynamic programming argument.

Optimality of Deterministic and Randomized LRT(s)

Theorem

It is **sufficient** to consider the boundary of $\mathcal{R}_{it|u_1^{t-1},u_2^{t-1}}$ and the vertices of sub-regions as the optimal candidates of $(p^{\mathsf{F}}_{it|u_1^{t-1},u_2^{t-1}}, p^{\mathsf{D}}_{it|u_1^{t-1},u_2^{t-1}})$.

Corollary:

If $\gamma^*_{it|u_1^{t-1},u_2^{t-1}}$ is not achieved by a **deterministic** LRT then can be realized by a **randomized** strategy of two LRTs.

Summary

- We proposed new a **privacy-per-design framework** for distributed detection problems:
  - Introduced *detection-theoretic privacy metrics*;
  - Formulated *privacy-constraint and privacy-aware problems*;
  - Identified *necessary and sufficient conditions* for optimal decision strategies
  - Studied parallel, serial, differential-privacy, and sequential setups
- It is possible to improve detection performance under maximal privacy constraint.
- Concept is interesting due to *low complexity* at remote sensors even with *many sensors* and therefore *low delay*.
  - We just started to explore the ideas...

Summary

- We proposed new a **privacy-per-design framework** for distributed detection problems:
  - Introduced *detection-theoretic privacy metrics*;
  - Formulated *privacy-constraint and privacy-aware problems*;
  - Identified *necessary and sufficient conditions* for optimal decision strategies
  - Studied parallel, serial, differential-privacy, and sequential setups
- It is possible to improve detection performance under maximal privacy constraint.
- Concept is interesting due to *low complexity* at remote sensors even with *many sensors* and therefore *low delay*.
  - We just started to explore the ideas...

Thank you for your attention!