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]\(^1\) Eavesdropper (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

---


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
  \[ \gamma_F(u_1, \ldots, u_N) = U_F, \]

- eavesdropper
  \[ \gamma_E(u_1) = U_E. \]

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 $c_{UF,H}(u_F, h)$.
  - Bayesian risk of the fusion node $c_F = \sum_{u_F, h} p_{U_F,H}(u_F, h)c_{U_F,H}(u_F, h)$

- **Neyman-Pearson approach**: Maximize detection probability
  $p^D_F = p_{U_F|H}(1|1)$ with an upper bound on the false alarm probability $p^E_F = p_{U_F|H}(1|0)$

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.


Privacy-Constrained Bayesian Detection Problem

- Bayesian approach:
  - Define costs for Eve $c_{U_E,H}(u_E, h)$
  - Assume Eve knows prior probability $p_H(h)$
- Privacy metric (minimal Bayesian risk, since Eve is greedy):

$$c_{E}^{\text{min}} = \min_{\gamma_E} c_E = \min_{\gamma_E} \sum_{u_E, h} p_{U_E,H}(u_E, h)c_{U_E,H}(u_E, h).$$

- A detection-theoretic operational privacy metric!

Privacy-constrained parallel distributed Bayesian detection problem

$$\min_{\gamma_1, \gamma_2, \cdots, \gamma_N, \gamma_F} c_F, \quad \text{s.t.} \quad c_{E}^{\text{min}} \geq \beta.$$
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 $\gamma_1$

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

while all other decision strategies are fixed.

Observations:

- Strategy $\gamma_1$ determines operation point $(p^F_1, p^D_1)$.
- **Objective** $c_F(p^F_1, p^D_1) = a_1 p^F_1 + b_1 p^D_1 + c_1$ is linear in $(p^F_1, p^D_1)$.
- **Constraints** $c_{E}^{\min} \geq \beta \iff c_{E}(p^F_1, p^D_1) \geq \beta, \forall \gamma_{E}$ are linear in $(p^F_1, p^D_1)$.  

Illustration of Privacy-Constrained PBPO

- A linear objective over a convex set:

Person-by-Person Optimality

It is **sufficient** to consider operating points \((p^F_1, p^D_1)\) 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.
Remark

The algorithmic method of PBPO\textsuperscript{6} can be easily extended to incorporate the privacy constraint.

---

Independent $N_i \sim \mathcal{N}(0, 1)$

Bayesian costs such that $c_F$ and $c_E^{\min}$ 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?
**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):**

  \[ \hat{p}_E^{D,\gamma} = \max_{\gamma E} p_E^D, \quad \text{s.t.} \quad p_E^F \leq \gamma. \]

Privacy-constrained Neyman-Pearson problem

\[
\max_{\gamma_1, \gamma_2, \gamma_F} p_F^D, \quad \text{s.t.} \quad p_F^E \leq \lambda, \hat{p}_E^{D,\gamma} \leq \delta.
\]
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 \( \gamma = \lambda \),

- \( p_D^{F} \) increases along line segments \( A \rightarrow B \) and \( A \rightarrow C \) 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.

---

4-ary hypothesis $H = (H_A, H_B)$
- 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

$$\min_{\gamma_1, \gamma_2, \ldots, \gamma_N, \gamma_A} c_A, \quad \text{s.t.} \quad c_B^{\min} \geq \beta.$$
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:**

\[ a_if_{Y_i|H_A,H_B}(y_i|0,0) + b_if_{Y_i|H_A,H_B}(y_i|1,0) + c_if_{Y_i|H_A,H_B}(y_i|0,1) + \\
\quad d_if_{Y_i|H_A,H_B}(y_i|1,1) \left\{ \begin{array}{ll}
  0 & \text{if } u_i=0 \\
  \geq & \text{if } u_i=1
\end{array} \right. \]
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 $T$

---

Privacy-Concerned Detection Problem

- Independently randomized local decision strategies:
  \[ \gamma_{it}(y_{it}, u_{1}^{t-1}, u_{2}^{t-1}) = U_{it} \]

- \( \gamma_{F}, \gamma_{E} \) are deterministic sequential detection strategies.

- **Privacy-concerned** Bayesian risk:
  \[ c_{P} = \alpha c_{F} - (1 - \alpha)c_{E}^{\text{min}}, \alpha \in [0, 1]. \]

- Privacy-concerned parallel distributed Bayesian sequential detection problem:
  \[ \min_{\gamma_{1}^{T}, \gamma_{2}^{T}, \gamma_{F}} c_{P}. \]
Privacy-Concerned Person-by-Person Optimality

$$\min_{k \in \{1, \ldots, K\}} \min_{(p_{it|u_1^{t-1},u_2^{t-1}}, p_{it|u_1^{t-1},u_2^{t-1}}) \in R_{itk|u_1^{t-1},u_2^{t-1}}} \alpha c_F - (1 - \alpha)c_{E_k},$$

convex optimization

with convex set

$$R_{itk|u_1^{t-1},u_2^{t-1}} = \left\{ (p_{it|u_1^{t-1},u_2^{t-1}}, p_{it|u_1^{t-1},u_2^{t-1}}) \Bigg| \exists \gamma_{it|u_1^{t-1},u_2^{t-1}} \text{ with } \begin{array}{c} c_{E_k} \leq c_{E1} \\ \vdots \\ c_{E_k} \leq c_{EK} \end{array} \right\},$$

and $c_F, c_{E_k}$ are linear functions of $p_{it|u_1^{t-1},u_2^{t-1}}, p_{it|u_1^{t-1},u_2^{t-1}}$. 
Illustration of Sub-Regions with 4 Candidates of $\gamma^*_E$

- When privacy-concerned person-by-person optimizing $\gamma_F$, use the dynamic programming argument.
Theorem

It is sufficient to consider the boundary of \( R_{it|u_{t-1}^{1},u_{t-1}^{2}} \) and the vertices of sub-regions as the optimal candidates of \((p_{it|u_{t-1}^{1},u_{t-1}^{2}}, p_{it|u_{t-1}^{1},u_{t-1}^{2}}))\).

Corollary:
If \( \gamma_{it|u_{t-1}^{1},u_{t-1}^{2}}^{*} \) is not achieved by a deterministic LRT then can be realized by a randomized strategy of two LRTs.
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!