NANYANG TECHNOLOGICAL UNIVERSITY
School of Electrical and Electronic Engineering

MULTILAYER SENSOR NETWORK FOR INFORMATION PRIVACY

XIN HE, WEE PENG TAY

MOTIVATIONS

With the ubiquitous adoption of Internet of Things (IoT) devices like on-body sensors, smart home appliances, and smart phones, massive amounts of data are being collected by service providers.

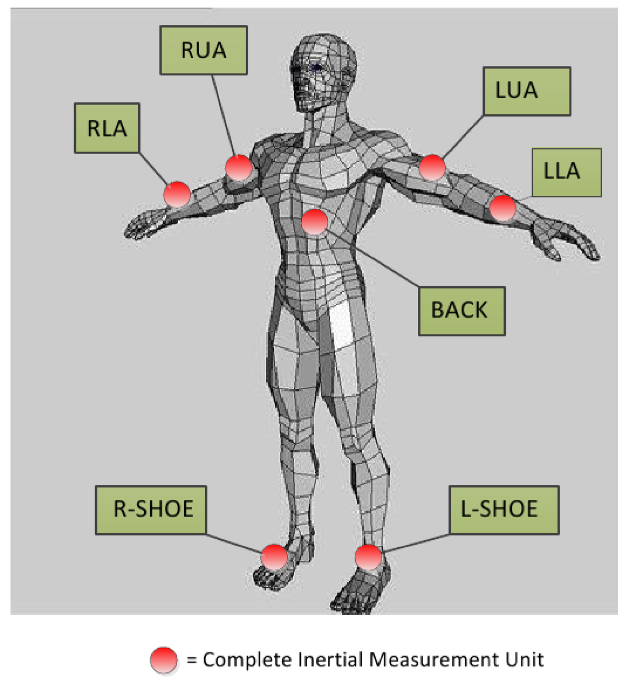

Fig. 1: Information privacy in on-body sensor network (UCI dataset).

The observed data can be exploited for

- Public information: exercise logs for health monitoring,
- Private information: users' private behaviors, habits, emotion and medical condition,

PROBLEM FORMULATION

With training samples $(\mathbf{X}_i, p_i, q_i)_{i=1}^{l}$, an optimization problem is proposed to find $\mathbf{G} = \{\mathbf{G}^m\}_{m=1}^{M}$ so as to

- minimize the regularized empirical risk of detecting public hypothesis p,
- keep the regularized empirical risk of detecting q above a given privacy threshold θ.

$$\min_{\mathbf{G}\in\mathcal{G},\mathbf{w}_\alpha} \frac{1}{l}\sum_{i=1}^{l}\phi(p_i\langle\mathbf{w}_\alpha,\Phi(\mathbf{Z}_i)\rangle_{\mathcal{H}}) + \frac{\lambda_\alpha}{2}\|\mathbf{w}_\alpha\|_2^2,$$

$$\text{s.t.} \min_{\mathbf{w}_\beta} \sum_{k\in\{-1,1\}} \frac{1}{2|\mathcal{S}_k|}\sum_{i\in\mathcal{S}_k}\phi(q_i\langle\mathbf{w}_\beta,\Phi(\mathbf{Z}_i)\rangle_{\mathcal{H}}) + \frac{\lambda_\beta}{2}\|\mathbf{w}_\beta\|_2^2 \geq \theta$$

where $\phi(\cdot)$ is a convex loss function, $\Phi(\cdot)$ is a feature map, $\mathcal{S}_k$ with $k\in\{-1,1\}$ contains the indexes of the training sample with label $q_i = \{-1,1\}$, respectively, $\mathbf{w}_\alpha$ and $\mathbf{w}_\beta$ are the fusion center decision rules for the public and private hypothesis, respectively.

METHODOLOGY

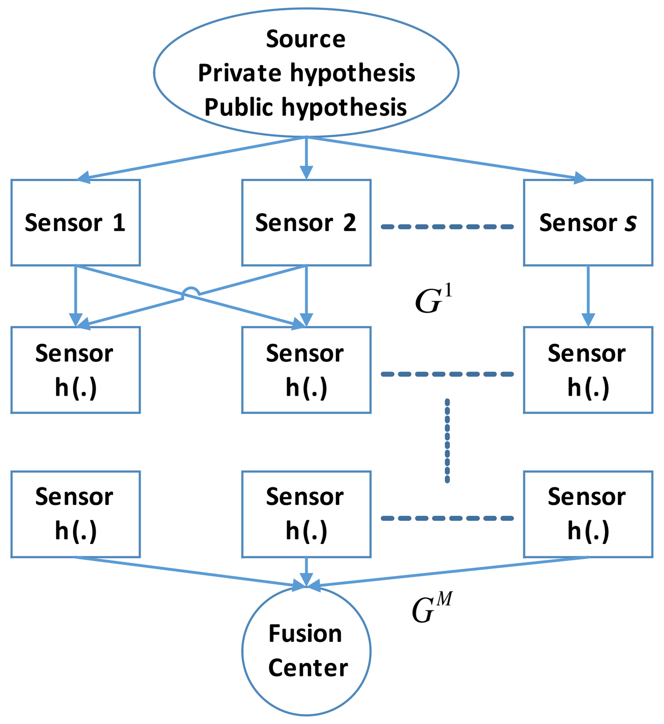

Fig. 2: Information privacy using a multilayer network.

After a repeated linear and nonlinear distortion of the observed data, the distorted data at the fusion center is

$$\mathbf{Z}(\mathbf{X}) = \mathbf{G}^M h(\mathbf{G}^{M-1}h(\cdots h(\mathbf{G}^1\mathbf{X}))).$$

The target of the weighting matrices $\{\mathbf{G}^m\}_{m=1}^{M}$ is to:

- Extract the public hypothesis related feature.
- Distort the privacy hypothesis related feature.

THE PROPOSED ALGORITHM

1. Finding the threshold θ:

   In the dual formulation, the best empirical risk of detecting the private hypothesis q under the worst case $\mathbf{G}$ is

$$\max_{\mathbf{G}\in\mathcal{G},\boldsymbol{\beta}} - \sum_{k\in\{-1,1\}}\frac{1}{2|\mathcal{S}_k|}\sum_{i\in\mathcal{S}_k}\phi^*(-2|\mathcal{S}_k|\beta_i) - \frac{1}{2\lambda_\beta}(\mathbf{q}\circ\boldsymbol{\beta})^T\mathbf{K}(\mathbf{G},\mathbf{X})(\mathbf{q}\circ\boldsymbol{\beta})$$

   Let the objective value be θ^*, then we chose $\theta = p\theta^*$, $p\in[0,1]$.

2. Optimizing the weighting matrices:

   (a) Without constraint on $\{\mathbf{G}^m\}_{m=1}^{M}$: gradient descent with line search such that the constraint is satisfied.

   (b) Positive semi-definite constraints on $\{\mathbf{G}^m\}_{m=1}^{M}$: modified mirror descent method to obtain a closed-form solution for gradient updating.

RESULT ON UCI ACTIVITY RECOGNITION

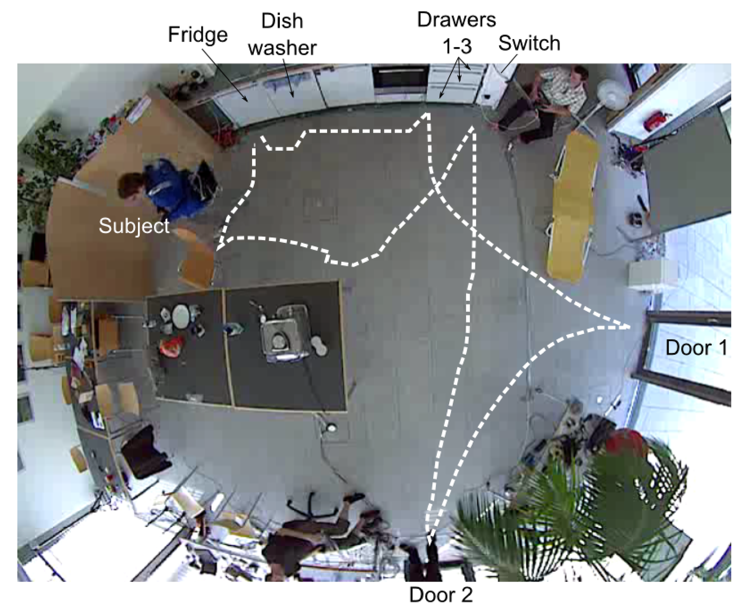

Fig. 3: The UCI OPPORTUNITY activity recording room.

Sensors	Inertial measurement
Public hypothesis	open or close a door ?
Private hypothesis	walking or standing ?

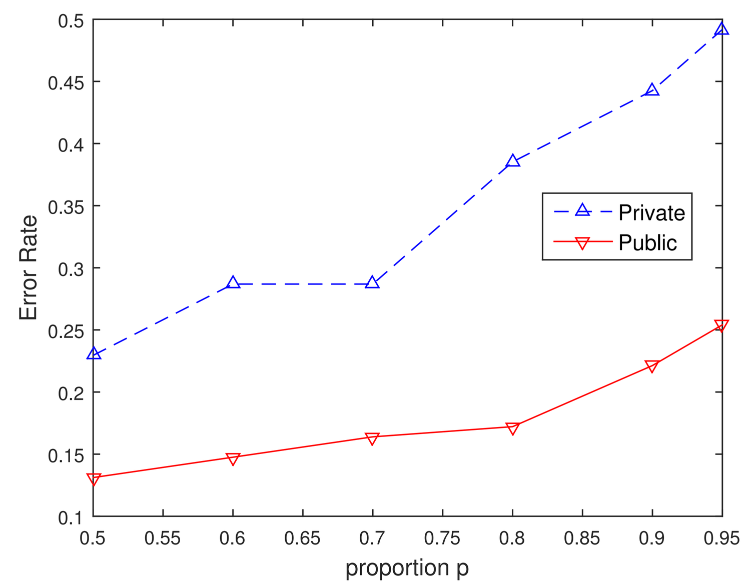

Fig. 4: The impact of the proportion p on public and private hypothesis test.

RESULT ON DALLAS ACTION RECOGNITION

Sensors	Inertial measurement+Kinetic camera
Public hypothesis	Boxing action existence ?
Private hypothesis	Baseball action existence ?

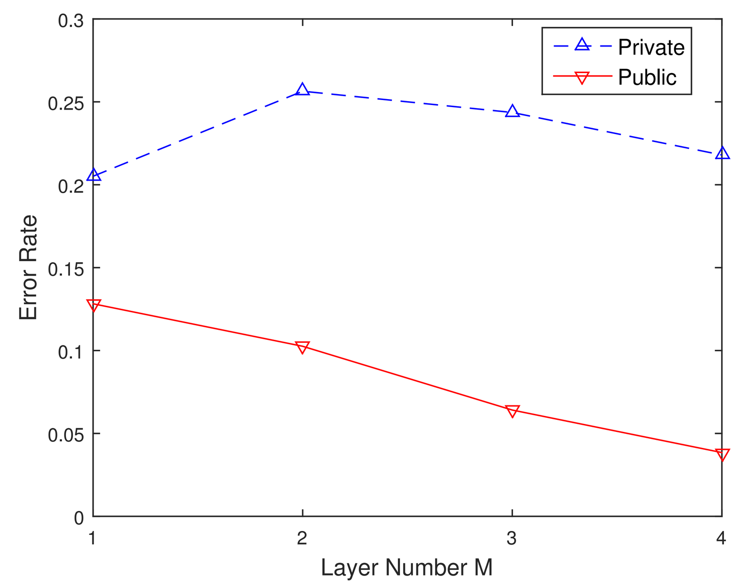

Fig. 5: Inertial sensor and kinetic camera.

- Note: 25% is the worst error rate for this four action setting.

REAL IMAGE EXPERIMENT

Fig. 6: Image experiment. The presence or absence of a gun and cash are the public and private hypothesis, respectively.

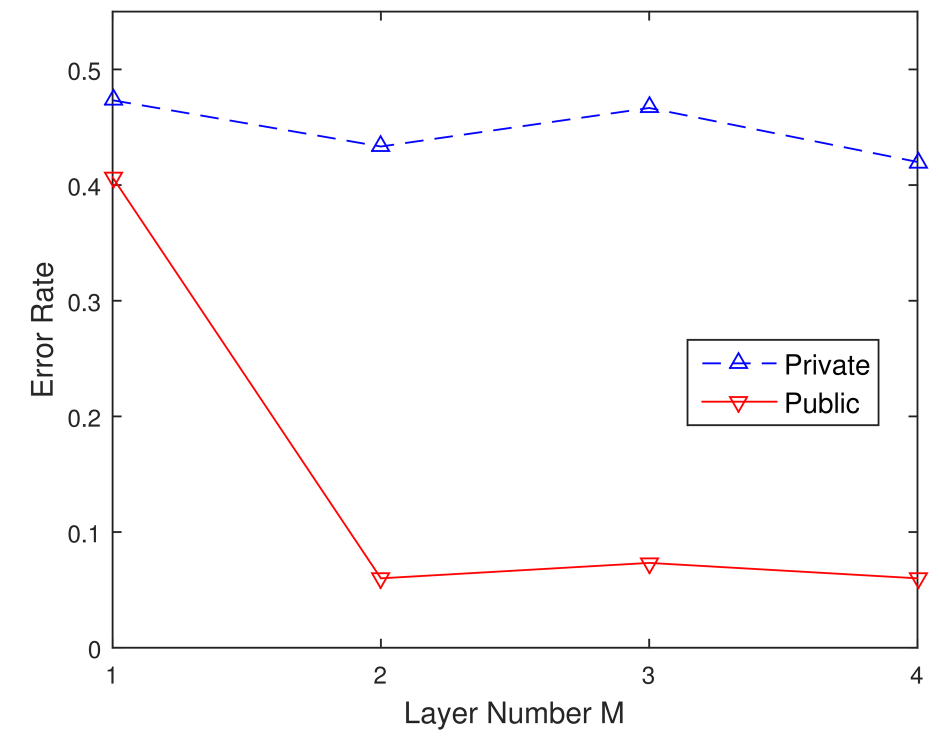

Fig. 7: One type of gun and cash.

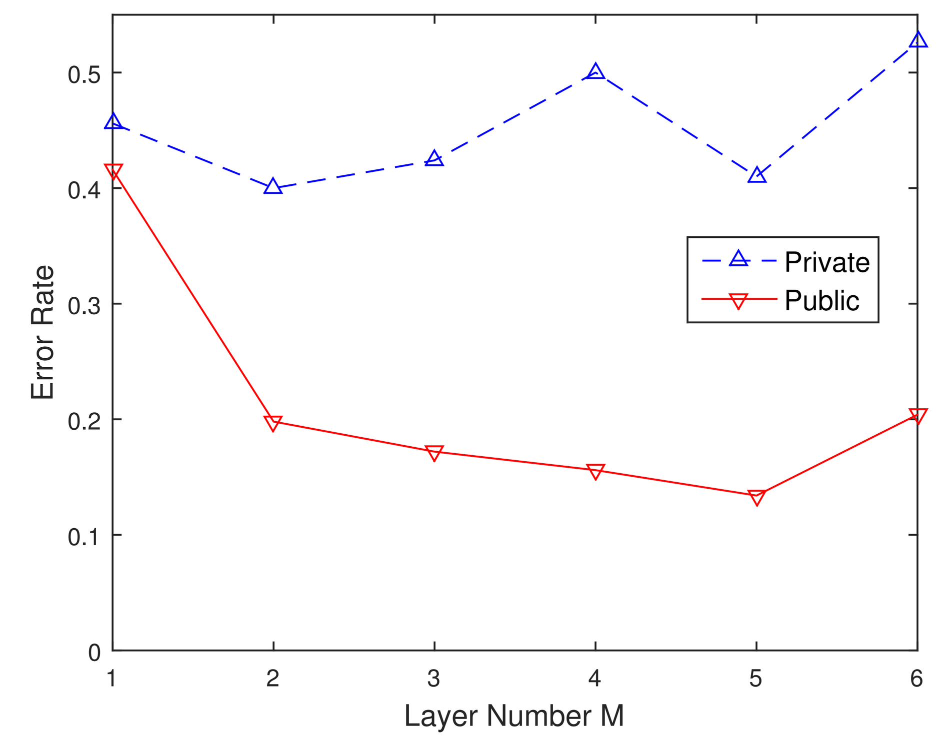

Fig. 8: Three types of gun and cash.

- Conclusion: A proper layer number M should be chosen to match the multilayer model and the dataset.