Privacy-Preserving Collaborative Deep Learning with Application to Human Activity Recognition

Conference Paper · November 2017

DOI: 10.1145/3132847.3132990

CITATIONS
23

READS
1,099

4 authors:

Lingjuan Lyu
University of Melbourne
42 PUBLICATIONS **286** CITATIONS
SEE PROFILE

Xuanli He
Monash University (Australia)
18 PUBLICATIONS **175** CITATIONS
SEE PROFILE

Yee Wei Law
University of South Australia
88 PUBLICATIONS **2,264** CITATIONS
SEE PROFILE

Marimuthu Palaniswami
University of Melbourne
595 PUBLICATIONS **20,366** CITATIONS
SEE PROFILE

Privacy-Preserving Collaborative Deep Learning with Application to Human Activity Recognition

Lingjuan Lyu
The University of Melbourne
Melbourne, Victoria 3010
llv@student.unimelb.edu.au

Xuanli He*
The University of Melbourne
Melbourne, Victoria 3010
xuanlih@student.unimelb.edu.au

Yee Wei Law
University of South Australia
Adelaide, South Australia 5095
yeewei.law@unisa.edu.au

Marimuthu Palaniswami
The University of Melbourne
Melbourne, Victoria 3010
palani@unimelb.edu.au

ABSTRACT

The proliferation of wearable devices has contributed to the emergence of *mobile crowdsensing*, which leverages the power of the crowd to collect and report data to a third party for large-scale sensing and collaborative learning. However, since the third party may not be honest, privacy poses a major concern. In this paper, we address this concern with a two-stage privacy-preserving scheme called *RG-RP*: the first stage is designed to mitigate *maximum a posteriori* (MAP) estimation attacks by perturbing each participant's data through a nonlinear function called *repeated Gompertz* (RG); while the second stage aims to maintain accuracy and reduce transmission energy by projecting high-dimensional data to a lower dimension, using a row-orthogonal *random projection* (RP) matrix. The proposed RG-RP scheme delivers better recovery resistance to MAP estimation attacks than most state-of-the-art techniques on both synthetic and real-world datasets. For collaborative learning, we proposed a novel LSTM-CNN model combining the merits of Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). Our experiments on two representative movement datasets captured by wearable sensors demonstrate that the proposed LSTM-CNN model outperforms standalone LSTM, CNN and Deep Belief Network. Together, RG+RP and LSTM-CNN provide a privacy-preserving collaborative learning framework that is both accurate and privacy-preserving.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**; • **Human-centered computing → Collaborative and social computing**; • **Computing methodologies** → *Artificial intelligence*; *Neural networks*;

KEYWORDS

Privacy-Preserving; Deep learning; Collaborative learning; Mobile crowdsensing

1 INTRODUCTION

The terms *participatory sensing* and *mobile crowdsensing* (MCS) refer to the new data crowdsourcing paradigm, where participants use ever more capable wearable devices or mobile phones to collect sensor readings and analyse them in the "cloud". MCS promises novel applications in healthcare, fitness, etc.

In healthcare, thanks to the proliferation of wearable devices for measuring heart rate, blood pressure, movement, and many other physiological parameters, patients and elderlies can now be monitored through their wearable devices, so that care can be given on time. Just as importantly, the wearable sensor data can be mined to reveal insights about the activities of the user community through *collaborative multi-user activity recognition* [16].

In fitness applications, integrated social networking drives data sharing, where there exists a large community whose members upload their wearable sensor data to online services that compile and publish community statistics. The ability to track and share fitness data online can enable positive reinforcement by peer groups, which encourages people to remain engaged in their physical activities and set new goals [28].

The trend, as we can see from the sample applications above, is that online services is expected to compute useful statistics from community data. While an online service is generally not malicious, it may be *semi-honest*, i.e., it follows the protocol honestly, but may attempt to learn or infer sensitive information from the users' data. In many domains, especially biomedicine, user data presents great privacy risks. Fitness data collected by wearable devices, for example, can include heart rate, location, calories consumption, stress level, and other data that can reveal the user's identity, ethnicity, disease risks, and other sensitive information [19]. Users are also concerned that their data might be sold to third parties without their consent. To remove the deterrents for users to share their data and benefit from the community knowledge discovery afforded by MCS, service providers need privacy-preserving algorithms that deliver a reasonable trade-off between data privacy and utility. Hence the problem we are addressing is *how a data owner can release its data*

*This author contributed equally to the first authorship.

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11…$15.00
https://doi.org/10.1145/3132847.3132990

with the guarantees that the original sensitive information cannot be reidentified while the analytic properties of the data are preserved.

There exist cryptographic techniques that allow data mining algorithms to be modified to process encrypted data; these techniques fall under the heading of *secure multiparty computation* (SMC) [9]. SMC approaches achieve a high level of privacy and accuracy, at the expense of high computational and communication overhead for the participants, thereby doing a disservice to attracting participation. Moreover, they require the cooperation of all participants throughout the whole process of building the data mining models, and hence suffer a lack of scalability.

Considering mobile users' hardware, energy and time constraints, a more practical alternative to SMC is *randomisation/perturbation*, where the participants perturb their data in a computationally efficient way before sending them to the cloud service. However, in the case of cloud service colluding with any subset of the participants, randomised data are subject to statistical data recovery attacks [15, 20, 31]. We tackle this problem by designing a randomisation-based collaborative learning scheme that is resistant to data recovery by colluding attackers.

In this paper, we consider the scenario where a third party performs deep learning on the union of data contributed by a large number of participants. Specifically, we apply LSTM-CNN, which has demonstrably good performance for dealing with time-series data. For privacy preservation, we propose a two-stage scheme called RG+RP, where in the first stage, each participant perturbs its data by passing the data through a nonlinear function called *repeated Gompertz* (RG) [26]; and in the second stage, each participant's data is projected to a lower dimension using a row-orthogonal *random projection* (RP) matrix [25]. The nonlinear function is designed to condition the probability density function (pdf) of the perturbed data to thwart *maximum a posteriori* (MAP) estimation attacks [31], whereas random projection compresses the data in a manner that stochastically preserves the Euclidean distances between pairs of data points as per the Johnson-Lindenstrauss Lemma [25].

We use two metrics to evaluate the effectiveness of our scheme. The first metric, ϵ*-recovery rate* [26], measures how much data can be recovered within a relative error of ϵ (see Sect. 5); this assesses how much privacy is preserved and how likely users' private data can be compromised. The second metric, recognition accuracy measures the effect of our privacy-preserving mechanism.

Our contributions are as follows: (i) The proposal of a two-stage perturbation mechanism for privacy-preserving collaborative learning, which maintains the privacy of both normal and anomalous records in terms of resistance to both MAP estimation attacks and *independent component analysis* (ICA) attacks. The random projection stage ensures the accuracy of time-series classification, especially for large data size [10]. To our knowledge, this is the first work in which a row-orthogonal matrix is used as the random projection matrix to be resistant to ICA attacks [25]. In addition, RP also saves transmission energy by compressing data. (ii) A novel LSTM-CNN model is applied to human activity recognition from wearable sensor data, that preserves learning accuracy even after the two-stage perturbation. Our solution shows great promise for building MCS applications that respect the participants' privacy.

2 RELATED WORK

This section reviews first the state of the art in deep learning for ubiquitous computing, and then randomisation schemes.

2.1 Deep learning in ubiquitous computing

In ubiquitous computing, wearable Human Activity Recognition (HAR) is attracting significant interest. HAR aims to recognise human activities via inertial, video and/or other sensors, enabling creation of ambient intelligence in smart environments, in which people of needs can be monitoried and timely cared for.

Movement data gathered from body-worn sensors are multivariate time-series data with inherent local dependency characteristics and relatively high spatial and temporal resolution. For analysing this kind of data, hidden Markov models (HMMs) have been widely used, owing to its capability of temporal pattern decoding. A HMM assigns probability values over a potentially infinite number of sequences. Since the probability values must sum to one, the distribution described by the HMM is constrained. Recently, deep learning has emerged as a family of learning models that aim to model high-level abstractions in data, and deep learning-empowered HAR is getting considerable attentions due to its power to learn deep structures of patterns [21]. Deep learning techniques have been shown to outperform the existing well-established methods that rely on hand-crafted feature extraction and shallow feature learning architectures, owing to its potential to uncover features that are tied to the dynamics of human motion production, from simple motion encoding in lower layers to more complex motion dynamics in upper layers, thus scaling up activity recognition to more complex activities [30].

Among various deep learning models, Deep Belief Networks (DBN) use Restricted Boltzmann Machines (RBMs) for learning, and avoid the local minimum problem with less training time. In comparison, Recurrent Neural Networks (RNN) can offer more discriminative power over DBN as they can encode/learn time sequential information. Given a series of input signals $x_1, \ldots, x_n$, their temporal information can be extracted by LSTM [18]

$$\begin{aligned} i_t &= \sigma(W_1 x_t + W_2 h_{t-1}), \\ \tilde{c}_t &= \tanh(W_3 x_t + W_4 h_{t-1}), \\ f_t &= \sigma(W_5 x_t + W_6 h_{t-1}), \\ o_t &= \sigma(W_7 x_t + W_8 h_{t-1}), \\ c_t &= c_{t-1} \odot f_t + i_t \odot \tilde{c}_t, \\ h_t &= c_t \odot o_t, \end{aligned} \tag{1}$$

where the subscript t denotes the time window, operator $\odot$ refers to component-wise multiplication, and σ is the logistic sigmoid function. h and c are the hidden state and the cell state respectively.

2.2 Privacy-preserving randomisation

Semantic privacy criteria are concerned with minimising the difference between adversarial prior and posterior knowledge about the individuals represented in a database. The standard semantic privacy criterion is *differential privacy*, which targets the scenario where a database server *answers queries* in a privacy-preserving manner by adding tailored exponential distributed noise to the query results [12]. In such a scenario, the database comprises private

data of *multiple individuals*. In the MCS scenario, participants are data owners who *publish data* (instead of answering queries) about *themselves alone*. Differential privacy can be made "distributed" to cater for the MSC scenario, but to prevent an attacker from filtering out the additive noise, a computationally intensive cryptographic protocol is needed to secure server-participants communications, as evidenced by the following schemes:

- Dwork et al.'s scheme [11] requires participants to communicate with each other (in addition to the cloud service), and use verifiable secret sharing to generate shares of random noise, imposing computational and communication costs on the order of 2^t, where t is the estimated maximum number of dishonest participants.
- Shi et al.'s scheme [33] enables participants to upload encrypted values to a data aggregator, who computes the sum of the encrypted values. These values are perturbed with geometric noise (which is approximately the discrete version of Laplace noise) to satisfy (ϵ, δ)-differential privacy, but the encryption relies on a *trusted dealer* allocating $q + 1$ secrets that sum to 0, to the data aggregator and the q participants.
- Ács et al.'s scheme [1] enables smart meters, organised into clusters, to send Laplace noise-tainted readings to an electricity distributor; but requires all meters in a cluster to share pairwise keys.

Most randomisation-based schemes use alternative privacy criteria. For our scheme, the criterion *recovery resistance* is defined in Sect. 5. This criterion is based on the *recovery rate* metric used in Sang et al.'s innovative study of attacks on randomisation-based schemes [32].

In general, randomisation-based techniques include (i) additive perturbation, (ii) multiplicative perturbation, (iii) geometric perturbation, and (iv) nonlinear transformation.

Additive perturbation: In 2000, Agrawal and Srikant [2] proposed the addition of independent and identically distributed (i.i.d.) noise to the original data, but additive noise can be filtered out in many cases [20]. One solution is to correlate the additive noise with the original data [14], but a participant can infer the data of another participant if their data are correlated [34].

Multiplicative perturbation: The original data is perturbed by multiplying it with some random noise and only the perturbed version is released for data analysis. Well-known multiplicative schemes include:

- *Rotation perturbation*: It relies on an orthogonal matrix with real entries whose columns and rows are orthogonal unit vectors [8]. This scheme is vulnerable to "known-input attacks" [15], where an attacker can recover the original data from its perturbed version with just a few leaked inputs.
- *Random projection* (abbreviated as RP): The original high-dimensional data can be projected onto a lower-dimensional random subspace by a random matrix whose columns have unit lengths [6]. The random matrix can be
  - An orthogonal matrix: such a matrix can be generated using the discrete cosine/Fourier transformation [29].
  - A stochastically orthogonal matrix: the elements of the matrix can be generated by sampling independent samples from the same zero-mean Gaussian distribution [25].

  When an n-attribute data matrix is multiplied by a $k \times n$ ($k < n$) RP matrix, the Euclidean distances and inner products between data points are approximately preserved thanks to the Johnson-Lindenstrauss Lemma, making the perturbed and compressed data suitable for distance-based privacy-preserving data mining [25]. However, if the original data follows a multivariate Gaussian distribution, a large portion of the data can be reconstructed via MAP estimation [24, 31]. In order to resist MAP attacks, nonlinear transformation can be introduced as the first stage before random projection to condition the pdf of the data points [13].
- *Uniform random transformation* (abbreviated as RT): The matrix can be a projection matrix whose elements are independently sampled from the uniform distribution $U(0, 1)$ [27]. It is shown that RT does not preserve the distance between data points, but does not distort the distance significantly for low-dimensional data, as shown in Proposition 1.

Geometric perturbation: This uses a mix of additive and multiplicative perturbations, where the data matrix $\mathbf{X}$ is mapped to $\mathbf{RX} + \mathbf{\Phi} + \mathbf{\Delta}$, where $\mathbf{R}$ is a rotation perturbation matrix, $\mathbf{\Phi}$ is a random translation matrix with identical entries, and $\mathbf{\Delta}$ is an i.i.d. Gaussian noise matrix [8]. It is known that without $\mathbf{\Delta}$, geometric perturbation is vulnerable to "known input attacks" [15], but there are no general results on how the $\mathbf{\Delta}$ term reduces the effectiveness of these attacks.

Nonlinear transformation: This method first appeared in conjunction with linear techniques to thwart Bayesian estimation attacks [5]. The general randomisation takes the form $\mathbf{B} + \mathbf{Q} \cdot \mathcal{N}(\mathbf{A} + \mathbf{RX})$, where $\mathbf{B}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{R}$ are random matrices, and $\mathcal{N}$ is a bounded nonlinear function [5]. The tanh function is found to preserve the distance between normal data points, but collapses the distance between outliers, making the function suitable for privacy-preserving anomaly detection [5], provided only the privacy of anomalous records needs to be protected.

Following Erfani et al. [13], Lyu et al. [26] proposed an improved two-stage perturbation scheme which relies on a nonlinear transformation and participant-specific $U(0, 1)$-distributed multiplication matrix to resist both MAP and ICA attacks. The improvements include the use of the *repeated Gompertz* function as the nonlinear transformation function for protecting both anomalous and normal records from Bayesian estimation attacks. However, a participant-specific perturbation matrix breaks the relationship between data points by adding an additional participant-specific noise to a uniform distributed matrix, which potentially limits its application to distance-based data analyses.

Most of the methods above do not achieve a good trade off between privacy and recognition accuracy. This paper shows how to successfully tailor the nonlinear transformation and random projection as privacy preserving mechanism for accurate recognition.

3 TECHNIQUE FOR MULTIPLICATIVE PERTURBATION

Euclidean distance-preserving data perturbation [8, 23] has been getting attention as it delivers a promising privacy/accuracy trade-off. Assume an organization owns a private, real-valued dataset $\mathbf{X}$ (row: feature, column: data record) and wishes to make it publicly

available for data analysis while keeping the individual records private. To accomplish this, $\mathbf{Y} = \mathbf{TX}$ is released to the public, where $\mathbf{T}$ preserves Euclidean distances between columns and is only known to the data owner. Therefore, many useful data mining algorithms, with only minor modification, can be applied to $\mathbf{Y}$ and produce exactly the same patterns that would be extracted from $\mathbf{X}$. Random projection preserves Euclidean distance and is especially suitable for high-dimensional big data analysis. The reduced dimensions are not a subset of the original dimensions but rather a transformation, which is relevant for privacy preservation.

For a more rigorous analysis, Let $\mathbf{T} = [t_{ij}]$, where $i = 1, \ldots, w$, $j = 1, \ldots, n$, and $w < n$. Given some distribution D, if $t_{ij} \sim D$ is independently and identically distributed (i.i.d.), we write $\mathbf{T} \sim D_{w\times n}$. In Proposition 1, we consider three types of distributions (and hence three types of $\mathbf{T}$), in terms of their ability to preserve Euclidean distances and inner products.

Proposition 1. *Suppose $\mathbf{T}$ is a $w \times n$ multiplicative perturbation matrix, where $w < n$. Then,*

- **Case 1** *If $\mathbf{T} \sim U_{w\times n}(0,1)$, where $U(0,1)$ is the uniform distribution on the interval $[0,1]$, then Euclidean distances and inner products between data points are not preserved, however, distances are approximately preserved for $w = n = 2$.*
- **Case 2** *If $\mathbf{T} \sim N_{w\times n}(0,\sigma_t^2)$, where $N_{w\times n}(0,\sigma_t^2)$ is the zero-mean, σ_t^2-variance Gaussian distribution, then Euclidean distances are preserved with a scaling factor, while inner products are exactly preserved.*
- **Case 3** *If $\mathbf{T} \sim O_{w\times n}(-1,1)$, where $O(-1,1)$ is the Gaussian orthogonal matrix distribution, then both Euclidean distances and inner products between points are exactly preserved.*

Proof. Let the distance vector between any two participants be $\boldsymbol{x} = [x_j], j = 1, \ldots, n$, then

$$\begin{aligned}
\mathsf{E}[\|\mathbf{T}\boldsymbol{x}\|_2^2] &= \mathsf{E}\left[\sum_{i=1}^{w}\left(\sum_{j=1}^{n} t_{ij}x_j\right)^2\right] \\
&= \sum_{i=1}^{w}\mathsf{E}\left[\sum_{j=1}^{n} t_{ij}^2x_j^2 + \sum_{\substack{k,l\in\{1,\ldots,n\}\\ k\neq l}} t_{ik}t_{il}x_kx_l\right] \\
&= \sum_{i=1}^{w}\sum_{j=1}^{n}\mathsf{E}[t_{ij}^2x_j^2] + \sum_{i=1}^{w}\sum_{\substack{k,l\in\{1,\ldots,n\}\\ k\neq l}}\mathsf{E}[t_{ik}t_{il}x_kx_l] \\
&= \sum_{i=1}^{w}\sum_{j=1}^{n}\mathsf{E}[t_{ij}^2]x_j^2 + \sum_{i=1}^{w}\sum_{\substack{k,l\in\{1,\ldots,n\}\\ k\neq l}}\mathsf{E}[t_{ik}]\,\mathsf{E}[t_{il}]x_kx_l.
\end{aligned} \tag{2}$$

Consider Eq. (2) in three cases:

- **Case 1** $t_{ij} \sim U(0,1) \implies \mathsf{E}[t_{ij}^2] = \frac{1}{3}$, and $\mathsf{E}[t_{ik}] = \mathsf{E}[t_{il}] = \frac{1}{2}$. Hence,

$$\mathsf{E}[\|\mathbf{T}\boldsymbol{x}\|_2^2] = \frac{w}{3}\|\boldsymbol{x}\|_2^2 + \frac{w}{4}\sum_{\substack{k,l\in\{1,\ldots,n\}\\ k\neq l}} x_kx_l. \tag{3}$$

  The new distance $\|\mathbf{T}\boldsymbol{x}\|_2^2$ comprises two parts in expectation: the first part is a scaled version of the original distance, whereas the second part is a scaled sum of products of the original distance components. So the Euclidean distance is not preserved.

  Now we show that a uniform distribution matrix does not distort distance too much for $w = n = 2$. According to Young's inequality, $x_kx_l \le \frac{x_k^2+x_l^2}{2}$. Without loss of generality, suppose $k = 1$, then

$$\sum_{l\in\{2,\ldots,n\}} x_1x_l \le \frac{(n-1)x_1^2}{2} + \frac{x_2^2 + \ldots + x_n^2}{2}.$$

  Thus, for any pair of $k, l \in \{1, \ldots, n\}$, $k \neq l$, the following inequality exists:

$$\sum_{\substack{k,l\in\{1,\ldots,n\}\\ k\neq l}} x_kx_l \le (n-1)(x_1^2 + x_2^2 + \ldots + x_n^2) = (n-1)\|\boldsymbol{x}\|_2^2.$$

  Applying the inequality above to Eq. (3),

$$\mathsf{E}[\|\mathbf{T}\boldsymbol{x}\|_2^2] \le \frac{w(3n+1)}{12}\|\boldsymbol{x}\|_2^2.$$

  When $n = 2$ and $w = 2$ (equidimensional projection), we have $\mathsf{E}[\|\mathbf{T}\boldsymbol{x}\|_2^2] \le \frac{14}{12}\|\boldsymbol{x}\|_2^2$, which implies that uniform random transformation approximately preserves Euclidean distances when a two-dimensional dataset is transformed in the two-dimensional space.

  In addition, we notice that a uniformly distributed transformation matrix does not preserve the inner product.
- **Case 2** $t_{ij} \sim N(0,\sigma_t^2) \implies \mathsf{E}[t_{ij}] = 0$, and $\mathsf{E}[\|\mathbf{T}\boldsymbol{x}\|_2^2] = \frac{w}{3}\|\boldsymbol{x}\|_2^2$, thus Euclidean distances between new data points are preserved in expectation with a scaling factor.

  As shown in [22, Lemma 5.1.2], $\mathsf{E}[\mathbf{T}^\top\mathbf{T}] = w\sigma_t^2\mathbf{I}$. Let $\mathbf{X}_{n\times m_1}$ represent m_1 data points in $\mathbb{R}^n$, and $\mathbf{Y}_{n\times m_2}$ represent m_2 data points in $\mathbb{R}^n$. The projected matrix for $\mathbf{X}$ and $\mathbf{Y}$ are: $\mathbf{U} = \frac{1}{\sqrt{w}\sigma_t}\mathbf{TX}$ and $\mathbf{V} = \frac{1}{\sqrt{w}\sigma_t}\mathbf{TY}$, and it is easy to check that $\mathsf{E}[\mathbf{U}^\top\mathbf{V}] = \mathsf{E}[\mathbf{X}^\top\mathbf{Y}]$, thus the inner product between data points is preserved.
- **Case 3** If $\mathbf{T}$ is a random orthogonal matrix, then

$$\|\mathbf{T}\boldsymbol{x}\|_2^2 = (\mathbf{T}\boldsymbol{x})^\top\mathbf{T}\boldsymbol{x} = \boldsymbol{x}^\top\mathbf{T}^\top\mathbf{T}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x},$$

  so both the Euclidean distance and inner product between points are preserved exactly.

□

Proposition 1 shows the orthogonal distribution preserves the Euclidean distance and inner product exactly. This is why orthogonal matrices are used for random projection in the subsequently proposed scheme.

We summarize the properties for different distributions of the multiplicative matrix in Table 1. Orthogonal transformation preserves both Euclidean distance and inner product without scaling, which is beneficial for guaranteeing recognition accuracy. And in order to resist "known-input attacks", which is inherent in orthogonal transformation, the first stage relies on a nonlinear function. For ICA attacks, we use row-orthogonal matrix (the rows of T still retain the orthogonal property, the norm of each row equals 1, and $\mathbf{T}^\top\mathbf{T} = \mathbf{I}$) as our random projection matrix in the second stage.

Table 1: Comparing multiplicative perturbation matrices

Distribution	Euclidean distance	Inner product	Scaling
$U_{w\times n}(0,1)$	×	×	NA
$N_{w\times n}(0,\sigma_t^2)$	✓	✓	✓
$O_{w\times n}(-1,1)$	✓	✓	×

4 THE RG+RP SCHEME

As depicted in Fig. 1, the general collaborative mobile crowd-sensing architecture comprises a set of participants $C = \{c_i : i = 1, \ldots, q\}$, a data mining cloud service $\mathcal{S}$, and a set of end-users $\mathcal{U}$. The cloud service is assumed to be semi-honest, i.e., it will never perform any malicious action to disrupt the protocols or compromise the participants but it might try to discover privacy-sensitive information of the participants, including colluding with some of the participants. Based on the state of the art in *privacy-preserving data mining* (PPDM), the following design criteria are considered:

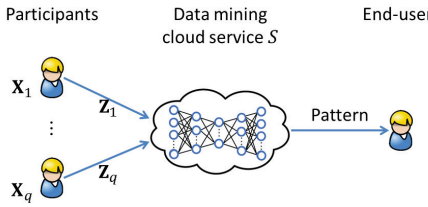


Figure 1: The mobile crowd-sensing architecture.

- **Resilience to Bayesian estimation attacks**: Bayesian estimation is a general attack that exploits the pdf of the original data. Gaussian data are particularly exploitable as the MAP estimation problem is reduced to a simple convex optimisation problem [31]. To prevent this reduction, a nonlinear transformation function is applied to the original data element-wise to condition the pdf.
- **Resilience to collusion**: The nonlinear function is chosen to be many-to-one, so that it does not have an inverse. This compounds the difficulty of extracting a participant's original data by colluding parties.
- **Resilience to ICA attacks**: Independent Component Analysis (ICA) aims to discover independent hidden factors that underlie a set of linear or nonlinear mixtures of some unknown variables. ICA attacks normally recover the original signals from only the observed mixture by designing a filter. To remove the prerequisites [22] for an ICA attack to succeed, we rely on random projection-based perturbation to enforce the reduced dimension to be lower than the half of the original dimension.

4.1 RG+RP

In the setup phase, each participant c_i is given a $w \times n$ ($w < n$) row-orthogonal random projection matrix, denoted $\mathbf{T}$, which is generated by QR-decomposing an $n \times n$ Gaussian distributed matrix $\sim N_{n\times n}(0,1)$. Suppose participant c_i is contributing data $\mathbf{X}_i \in \mathbb{R}^{n\times m_i}$ to the cloud service S for deep learning. The participant transforms $\mathbf{X}_i$ to $\mathbf{Z}_i \in \mathbb{R}^{w\times m_i}$ in two stages:

Stage 1: The participant transforms $\mathbf{X}_i$ to $\mathbf{Y}_i$, by applying the nonlinear perturbation function $\mathcal{N}$ element-wise:

$$\mathbf{Y}_i = \mathcal{N}(\mathbf{X}_i). \tag{4}$$

$\mathcal{N}$ is chosen to be the repeated Gompertz function:

$$\begin{aligned} N(x) \stackrel{\text{def}}{=} & a_1 e^{-b_1 e^{-c_1 x - d_1}} u(0.35 - x) \\ & + \left(0.5 + a_2 e^{-b_2 e^{-c_2 x - d_2}}\right) u(x - 0.35), \end{aligned} \tag{5}$$

where the parameters $a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2$ are defined in Fig. 2, and $u()$ is the Heaviside step function. The derivation of the function parameters is explained in Sect. 5. Fig. 2 plots different nonlinear perturbation functions for comparison.

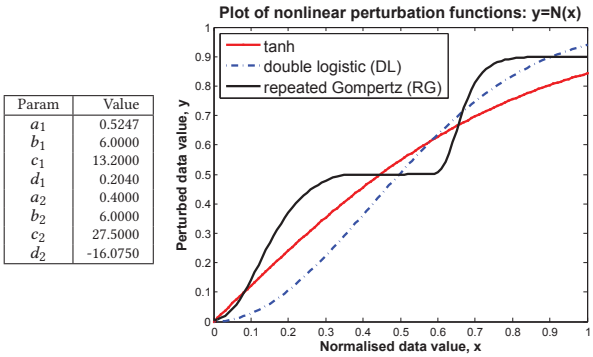


Figure 2: Parameters of repeated Gompertz function, and plot of different nonlinear perturbation functions. The tanh **function is $\mathcal{N}(x) = \tanh(\beta_t x)$, where $\beta_t \approx 1.23$ [5]. The double logistic function is $\mathcal{N}(x) = 1 - \exp(-\beta_{dl} x^2)$, where $\beta_{dl} \approx 2.81$ [13]. The repeated Gompertz function is defined in Eq.** (5).

Stage 2: Using the random projection matrix $\mathbf{T}$ initialized in the setup phase, the participant transforms $\mathbf{Y}_i$ to $\mathbf{Z}_i$ as follows:

$$\mathbf{Z}_i = \mathbf{T}\mathbf{Y}_i. \tag{6}$$

The participant then sends $\mathbf{Z}_i$ to the cloud service $\mathcal{S}$. Once $\mathcal{S}$ receives all the perturbed datasets $\mathbf{Z}_i$, $i = 1, \ldots, q$, it concatenates them as: $\mathbf{Z}_{\text{all}} = [\mathbf{Z}_1 | \cdots | \mathbf{Z}_q]$, then builds a deep learning model on $\mathbf{Z}_{\text{all}}$. The pseudocode for the collaborative deep learning procedure is shown in Algorithm 1. The recognition results can be used by end-users for further analysis or to provide valuable feedback to the participants.

5 PRIVACY ANALYSIS

Linear multiplicative perturbation schemes project a data vector (and hence the whole data matrix) to a lower dimensional space. Against such a scheme, an attacker has only an ill-posed problem in the form of an underdetermined system of linear equations $\mathbf{T}\boldsymbol{y} = \boldsymbol{z}$ to tackle with, where $\boldsymbol{z}$ is a projection of vector $\boldsymbol{y}$. In an underdetermined system, there exists no exact solution for $\boldsymbol{y}$, but provided sufficient prior information about $\boldsymbol{y}$, an approximation of the true $\boldsymbol{y}$ may be attainable. We characterise an attack by the extent of prior information available to the attacker.

Algorithm 1 The collaborative deep learning procedure

Role: Participant
$\mathbf{T} \leftarrow$ a $w \times n$ row-orthogonal matrix produced by QR-decomposing a Gaussian distributed matrix $\sim N_{n\times n}(0, 1)$
$\mathbf{Z}_i \leftarrow \mathbf{T}(\mathcal{N}(\mathbf{X}_i))$
Send $\mathbf{Z}_i$ to the cloud service

Role: Cloud service
Receives $\mathbf{Z}_i$, $i = 1, \ldots, q$
$\mathbf{Z}_{\text{all}} \leftarrow \left[\mathbf{Z}_1 | \mathbf{Z}_2 | \cdots | \mathbf{Z}_q\right]$
Build deep learning model on $\mathbf{Z}_{\text{all}}$
Send recognition results to the end-users

Role: End-user
Receive recognition results from the cloud server

In a *known input-output attack*, the attacker has some samples of the original data and all samples of the perturbed data, and knows the corresponding relationship between input sample and output sample [15]. In the mobile crowd-sensing scenario, the cloud service may collude with one or more participants to unravel other participants' data, in this scenario, the known input-output attack becomes an immediate concern. In the following, our privacy analysis targets a known input-output attack based on *maximum a priori estimation* (MAP) estimation. MAP estimation is based on Bayesian statistics, and is more general than maximum likelihood estimation because it takes the prior distribution into account.

This paper uses *recovery rate* [26] to measure the effectiveness of the reference attack. If the recovered version for a data vector $\boldsymbol{x}$ is $\hat{\boldsymbol{x}}$, then the *relative error* is defined as $\xi \stackrel{\text{def}}{=} \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2$, where $\|\cdot\|_2$ is the Euclidean norm. Denote the joint distribution of ξ and $\boldsymbol{x}$ by $p_{\Xi,\boldsymbol{X}}(\xi, \boldsymbol{x})$, then the *$\epsilon$-recovery rate* with respect to the perturbation algorithm and attack is defined as

$$r_\epsilon(\mathcal{A}, p_D) \stackrel{\text{def}}{=} \int_{\xi=0}^{\epsilon} \int_{\boldsymbol{x}\in D_{\boldsymbol{x}}} p_{\Xi,\boldsymbol{X}}(\xi, \boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\xi, \tag{7}$$

where $D_{\boldsymbol{x}}$ is the domain of the data vector, and $\boldsymbol{x}$ is normalised. The joint distribution $p_{\Xi,\boldsymbol{X}}$ depends on the attack $\mathcal{A}$ and data distribution p_D. In the absence of an analytical expression for Eq. (7), the recovery rate is estimated as the fraction of test data that can be recovered to within a relative error of ϵ. At this point, we state the privacy definition formally as follows.

A probabilistic algorithm that takes p_D-distributed $\boldsymbol{x} \in \mathbb{R}^n$ as input and produces $\boldsymbol{z} \in \mathbb{R}^w$ as output is (ϵ, δ)-*recovery resistant* with respect to p_D and attack algorithm $\mathcal{A}$ if $r_\epsilon(\mathcal{A}, p_D) = \delta$.

Suppose the attacker is targeting a particular participant by trying to solve $\mathbf{Z} = \mathbf{TY}$ for $\mathbf{Y}$. We consider two scenarios of prior knowledge about the random matrix: where $\mathbf{T}$ is known, and where $\mathbf{T}$ is unknown.

5.1 Random matrix T is Known

In this worst-case scenario, the random projection matrix $\mathbf{T}$ is known exactly to the attacker, for example, when the attacker manages to predict the output of the victim's improperly initialised pseudorandom number generator (in fact, such a vulnerability was discovered on the Android mobile platform in mid-2013). We denote a column of $\mathbf{Z}$ and $\mathbf{Y}$ as $\boldsymbol{z}$ and $\boldsymbol{y}$ respectively. The MAP estimate of $\boldsymbol{y}$, given $\mathbf{T}$ and $\boldsymbol{z}$, is

$$\begin{aligned}\hat{\boldsymbol{y}} &= \arg\max_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{z}, \mathbf{T}) = \arg\max_{\boldsymbol{y}} \frac{p(\boldsymbol{z}|\mathbf{T}, \boldsymbol{y})p(\mathbf{T})p(\boldsymbol{y})}{p(\boldsymbol{z}|\mathbf{T})p(\mathbf{T})} \\ &= \arg\max_{\boldsymbol{y}\in\mathcal{Y}} \frac{p(\boldsymbol{y})}{\int_{\mathbb{R}^n} p(\boldsymbol{z}|\mathbf{T}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{y}} = \arg\max_{\boldsymbol{y}\in\mathcal{Y}} p(\boldsymbol{y}),\end{aligned} \tag{8}$$

where $\mathcal{Y} = \{\boldsymbol{y} : \boldsymbol{z} = \mathbf{T}\boldsymbol{y}\}$. Note:

- The factor $p(\boldsymbol{z}|\mathbf{T}, \boldsymbol{y})$ translates to the constraint $\boldsymbol{y} \in \mathcal{Y}$.
- The integral in the denominator does not contribute towards maximising $\boldsymbol{y}$.

If $\boldsymbol{y}$ is n-variate Gaussian with a positive definite covariance matrix, then Eq. (8) becomes an easily solvable quadratic programming problem [31, Theorem 1]. In oder to deter the accurate solution of Eq. (8), we rely on a nonlinear function $\mathcal{N}$ to transform the potentially Gaussian data distribution to a different distribution.

In the following, a novel nonlinear function "repeated Gompertz" is proposed and defined in Eq. (5). The traditional Gompertz function takes the standard form:

$$\text{Gompertz}(x) = ae^{-be^{-cx}}, \tag{9}$$

where the parameter a specifies the upper asymptote, b controls the displacement along the x axis, and c adjusts the growth rate of the function. A detailed explanation of how the proposed function is derived is given below. As $\tanh(\beta_t x)$ is good for protecting anomalous data points, the slope of the repeated Gompertz function is approximated as those of $\tanh(\beta_t x)$ at $x = 0$ and $x = 1$. The repeated Gompertz function is also designed to have a flat middle section so that the function cannot be inverted for that section, thus normal data points are protected. Through extensive experimentation, we found the geometry of the function in Fig. 2 to be good for protecting both anomalous and normal data points: (i) a Gompertz curve presenting a steep slope over the interval $[0, 0.35]$; and (ii) another Gompertz curve presenting a plateau over the interval $[0.35, 0.6]$, a steeper slope over the interval $[0.6, 0.75]$ and another plateau over the interval $[0.75, 1]$. The parameters of the two Gompertz functions are given in Fig. 2. This compositional structure inspired the name "repeated Gompertz".

5.2 Random matrix T is unknown

In the case where the attacker knows neither $\mathbf{T}$ nor $\mathbf{Y}$, the MAP estimates of $\mathbf{Y}$ and $\mathbf{T}$, given $\mathbf{Z}$, are

$$\begin{aligned}(\hat{\mathbf{T}}, \hat{\mathbf{Y}}) &= \arg\max_{\mathbf{T},\mathbf{Y}} p(\mathbf{T}, \mathbf{Y}|\mathbf{Z}) \\ &= \arg\max_{\mathbf{T},\mathbf{Y}} \frac{p(\mathbf{Z}|\mathbf{T}, \mathbf{Y})p(\mathbf{T})p(\mathbf{Y})}{\int\int p(\mathbf{Z}|\mathbf{T}, \mathbf{Y})p(\mathbf{T})p(\mathbf{Y})d\mathbf{T}d\mathbf{Y}} \\ &= \arg\max_{(\mathbf{T},\mathbf{Y})\in\Theta} p(\mathbf{T})p(\mathbf{Y}),\end{aligned} \tag{10}$$

where $\Theta = \{(\mathbf{T}, \mathbf{Y}) : \mathbf{Z} = \mathbf{TY}\}$. In a known input-output attack, $p(\mathbf{T})$ and $p(\mathbf{Y})$ are estimated as inputs to Eq. (10). Eq. (10) is a nonconvex optimisation problem that is harder to solve than Eq. (8). The repeated Gompertz is designed to make data recovery via Eq. (8) difficult when $\mathbf{T}$ is known. Now that $\mathbf{T}$ is unknown, the attacker is expected to get an even lower recovery rate by solving Eq. (10), which is a more difficult problem.

5.3 Underdetermined Independent Component Analysis (UICA)

As a statistical technique, ICA represents a set of random variables as linear combinations of statistically independent component variables. The aim of an ICA attack is to design a filter that can recover the original signals from only the observed mixture. ICA can separate out $\mathbf{T}$ and $\mathbf{Y}$, knowing only their product $\mathbf{Z} = \mathbf{TY}$, provided (i) The number of observed attributes must be at least as large as the independent attributes, $w \geq n$; (ii) the attributes are independent; (iii) at most one of the attributes is Gaussian; (iv) $\mathbf{T}$ must have full column rank. To resist an ICA attack, we enforce $w < n$, namely projecting data to a lower-dimensional subspace to make the problem of ICA underdetermined. In this case, even if the perturbation matrix $\mathbf{T}$ is known, the independent components cannot be obtained. Moreover, as shown in [25], if $w \leq (n+1)/2$, no linear filter can separate out the observed mixture $\mathbf{Z}$. It is demonstrated that an ICA attack cannot effectively breach the privacy of participants after random projection-based perturbation.

5.4 Privacy evaluation

This section presents the simulation and evaluation results of RG+RP in terms of its recovery resistance. Since RG+RP is resistant to ICA, the empirical experiments focus on obtaining the recovery rates of MAP estimation attacks against RG+RP, and comparing them to those of prior work, i.e., [25], [5] and [13]. Experiments are conducted on both synthetic and real datasets shown in Table 2.

Table 2: Datasets for evaluating recovery resistance

Datasets	#records (m)	Upspace dimension (n)	Downspace dimension (w)
Purely Gaussian	5000	15	8
Purely Laplace	5000	15	8
Abalone	4177	8	4
Forest	581012	54	27
Adult	48842	123	62
Gas	13910	128	64
OAR	77597	110	55
DSA	7500	315	158
HAR	7352	561	281
Smiley	20000	20	10
GME	101000	100	50

In order to evaluate the recovery resistance of RG+RP against MAP estimation attacks, experimental results are provided in terms of the ϵ-recovery rate defined in Eq. (7). In the absence of an analytical expression for Eq. (7), we estimate the ϵ-recovery rate as the fraction of test data that can be recovered to within a relative error of ϵ:

$$\hat{r}_\epsilon(\mathcal{A}, p_D) \stackrel{\text{def}}{=} \frac{\#\left\{\hat{\boldsymbol{x}}_i : \frac{\|\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2}{\|\boldsymbol{x}_i\|_2} \leq \epsilon, i = 1, \ldots, m\right\}}{m}, \quad (11)$$

where $\boldsymbol{x}_i$ and $\hat{\boldsymbol{x}}_i$ are the ith original data record and its attacker-estimated value respectively.

To execute MAP estimation, the attacker can either apply the formula in [31, Theorem 1], provided the original data is multivariate Gaussian distributed; or solve the constrained optimisation problem (8). To solve optimisation problem (8), the attacker needs to evaluate an objective function that is the pdf of the original data, which can be estimated from the leaked input samples using multivariate *kernel density estimation* (KDE). For KDE, we rely on Ihler and Mandel's Kernel Density Estimation Toolbox for MATLAB[1]. Among the kernels supported, we use the Epanechnikov kernel — which is optimal in the sense of the asymptotic mean integrated squared error — with uniform weights.

Table 3: Evaluated schemes

Scheme	Nonlinear perturbation function (stage 1)	Linear projection matrix (stage 2)
RP [25]	none	$\mathbf{T} \sim N_{w\times n}(0, 4)$
tanh+RP	tanh [5]	$\mathbf{T} \sim N_{w\times n}(0, 1)$
DL+RT [13]	double logistic	$\mathbf{T} \sim U_{w\times n}(0, 1)$
RG+RP	repeated Gompertz	$\mathbf{T} \sim O_{w\times n}(-1, 1)$

The schemes shown in Table 3 are evaluated in the *worst-case* scenario where the attacker knows exactly the victim's perturbation matrix. The results for different datasets are as follows.

Purely Gaussian datasets: For Gaussian datasets, Fig. 3 (a) shows that RG+RP provides significantly higher recovery resistance for both normal and anomalous data compared to the other schemes, except for the 0.2-recovery rate for the anomalous data case, which is slightly less effective than RP. An intuitive explanation for why RP excels at protecting anomalous data is that RP has the effect of "normalising" anomalous data in the range space of $\mathbf{T}$.

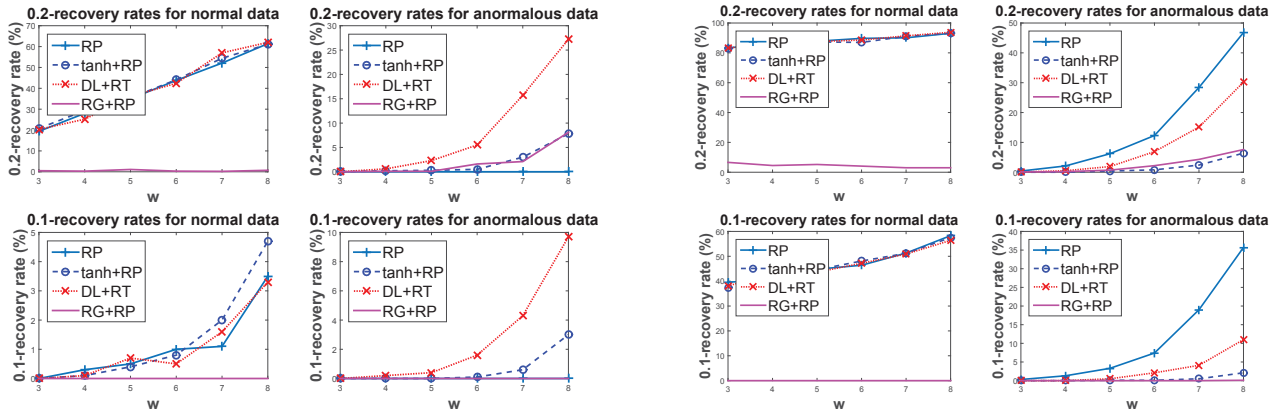
Purely Laplace datasets: Fig. 3 (b) demonstrates that RG+RP significantly outperforms other methods for Laplace datasets, and it is especially evident for normal data, except for the 0.2-recovery rate for the anomalous data case, which is slightly less effective than tanh+RP. Furthermore, the 0.1-recovery rate against RG+RP is below 10%, which is significantly lower than other schemes.

Assorted real and synthetic datasets: Consistent with the results for purely Gaussian and purely Laplace datasets, as shown in Fig. 4, RG+RP also outperforms tanh+RP and DL+RT in terms of recovery resistance for both normal data and anomalous data. It delivers low recovery rates in many cases, especially for DSA and HAR datasets, RG+RP achieves (0.1, 0)-recovery resistance.

6 PROPOSED LSTM-CNN AND PERFORMANCE EVALUATION

In deep learning models, CNNs and LSTMs are complementary in their modeling capabilities, as CNNs aim at reducing frequency variations, while LSTMs are especially good for temporal modeling. It is explored in [17] that RNNs outperform CNNs significantly on activities that are short in duration but have a natural ordering, however, CNNs is more suitable for prolonged and repetitive activities like walking or running. Therefore, it would be beneficial to exploit the

[1] http://www.ics.uci.edu/~ihler/code/kde.html



(a) Gaussian-distributed data.

(b) Laplace-distributed data.

Figure 3: Recovery rates of MAP estimation attacks against evaluated schemes, on $w \times 1000$ data projected from 15×1000 normalised Gaussian-distributed data (zero mean, identity covariance matrix) and Laplace-distributed data (zero mean, unity scale).

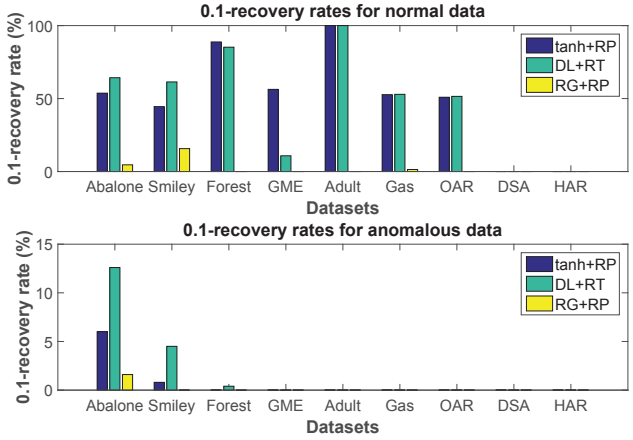


Figure 4: 0.1-recovery rates of MAP estimation attack against the evaluated schemes, on various datasets. The rank of the perturbation matrix, $w = \lfloor (n+1)/2 \rfloor$, where n is the number of features. Note zero recovery rates in many cases.

synergy of CNNs and LSTMs by combining them into one unified architecture and train them jointly. In our architecture, we leverage the local and dense property from convolution operation and learn the temporal structure by storing information in LSTM units by putting a CNN layer above LSTM layer, as illustrated in Fig. 5. First, the input features are fed into LSTM layers to reduce temporal variation. At each time step, the LSTM is capable of combining the previous information with the current input. Afterwards, we concatenate these outputs to a matrix $\mathbf{A} \in \mathbb{R}^{k \times s}$, where s denotes the number of time steps and k denotes the number of features.

Motivated by Blunsom et al. [7], we introduce convolutional layers applying one-dimensional filters across each column of time signals on the matrix $\mathbf{A} \in \mathbb{R}^{k \times s}$. We believe convolving the same filter can extract the dominant information over s consecutive time signals. In order to capture different positional information, multiple size-varied filters are utilized. Each convolution operation involves a filter $\mathbf{w} \in \mathbb{R}^{k \times m}$, which is applied to a window of m columns to produce a new feature. For example, a feature c_i can be generated from Eq. (12):

$$c_i = \text{ReLU}(\langle \mathbf{w}, \mathbf{A}_{i:i+m-1} \rangle + b), \tag{12}$$

where

- ReLU is the Rectified Linear Unit activation function;
- $\mathbf{A}_{i:i+m-1}$ refers to $\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_{i+m-1}$, for $\boldsymbol{x}_i \in \mathbb{R}^k$, $i = 1, \ldots, s-m+1$;
- $\langle \mathbf{A}, \mathbf{B} \rangle$ denotes the Frobenius inner product of $\mathbf{A}$ and $\mathbf{B}$;
- $b \in \mathbb{R}$ is a bias term.

Accordingly, for each filter, we can obtain a sequence of features: $c_1, c_2, \ldots, c_{s-m+1}$.

Like other CNN structures, the convolutional layer is followed by a maxpooling layer, which aims at returning the most crucial feature of a specific feature mapping, as indicated in Eq. (13):

$$c_{max} = \max(c_1, c_2, \ldots, c_{s-m+1}). \tag{13}$$

Once we obtain all predominant features from CNN, multi-layer perceptron (MLP) can be used to conduct activity classification.

6.1 Wearable Dataset

The initial accuracy evaluation of the proposed scheme is based on two real-world wearable datasets from the UCI Machine Learning Repository. The first wearable dataset is Human Activity Recognition database [3], which consists of recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. Each person performed six activities by wearing a smartphone (Samsung Galaxy S II) on the waist. From the embedded accelerometer and gyroscope, 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The labels were recorded by video. The sensor signals were pre-processed by applying noise filters and

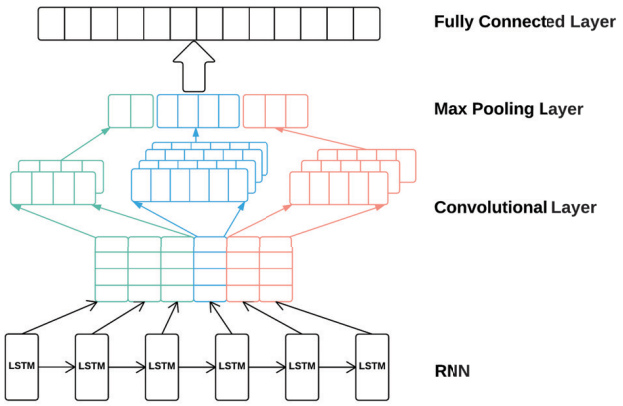


Figure 5: LSTM-CNN architecture. The bottom layer is a RNN nets with 6 time steps. At convolutional layer, we have 4 unigram filters, and 3 bigram filters and 2 trigram filters. Then maxpooling layer concatenates all maximum features of feature maps. Finally these features are fed into a MLP to perform activity classification.

then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). From each window, a vector of 561 features was obtained by calculating variables from the time and frequency domain.

Another wearable dataset is the Mobile Health (MH) dataset [4], which comprises body motion and vital signs recordings for ten volunteers while performing 12 common activities. Sensors placed on the subject's chest, right wrist and left ankle were used to measure the motion experienced by diverse body parts, namely, acceleration, rate of turn and magnetic field orientation. The sensor positioned on the chest also provides 2-lead ECG measurements, which can be potentially used for basic heart monitoring, checking for various arrhythmias or looking at the effects of exercise on the ECG. All sensing modalities were recorded using a video camera at a sampling rate of 50 Hz, and sampled in fixed-width sliding windows of 128 readings/window.

Both HAR dataset and MH dataset have been randomly partitioned into three sets, 70% training data, 15% validation data and 15% test data. The epoch that shows the best validation-set performance is used for testing purpose.

6.2 Model training and evaluation

For accuracy evaluation, we first establish baseline models, namely CNN, LSTM and DBN. The recognition accuracy for the raw data and the two-stage perturbed data are referred to as *Raw* and *Perturbed* respectively. The corresponding accuracy results for the HAR and MH datasets are shown in Table 4.

For the HAR dataset, we find that the LSTM-CNN architecture provides an 7.77% relative improvement over the LSTM, 6.71% relative improvement over the CNN, and 3.23% relative improvement over the DBN. We also notice that the proposed privacy-preserving two-stage mechanism lowers the accuracy by only 4.76% compared to the result from the raw data, and it achieves 93.75% accuracy,

Table 4: Comparing the proposed LSTM-CNN model with three baseline models in terms of recognition accuracy, using the HAR and MH datasets. The best results are highlighted in bold.

Model	HAR		MH	
	Raw	Perturbed	Raw	Perturbed
CNN	0.9225	0.9033	0.8399	0.8068
LSTM	0.9134	0.8048	0.8233	0.7996
DBN	0.9536	0.8455	0.8815	0.8237
LSTM-CNN	**0.9844**	**0.9375**	**0.9556**	**0.9208**

which is even higher than the accuracy delivered by the LSTM and CNN model trained on the raw data.

For the MH dataset, the LSTM-CNN architecture exhibits a similar accuracy trend, delivering an accuracy higher than 92% for both the raw data and perturbed data. Considering the impressive privacy benefits provided by our architecture, the accuracy is competitive. A plausible explanation for the comparable accuracy is that RP has no impact on accuracy degradation, and RG is designed to be nearly consistent in trend with the original data, thus causing negligible performance degradation. In addition, unlike simple linear projections, which have limited representational power, nonlinear projection can make data closer to linearly separable and thus easier to classify. Table 5 shows the results of varying LSTM depth. We found deeper models outperformed shallower ones by a small margin, while doubled the training time. Similarly, larger hidden units do not significantly improve accuracy.

7 CONCLUSIONS AND FUTURE WORK

In this paper, a two-stage randomisation-based scheme is presented for privacy-preserving collaborative deep learning. The scheme, called RG+RP, perturbs data in two stages: the first, nonlinear stage thwarts Bayesian estimation attacks, whereas the second, random projection resists independent component analysis attacks.

The nonlinear function is designed to condition the pdf of the perturbed data to protect both anomalous and normal data. Preliminary analysis and empirical evaluation indicate that compared with other non-linear schemes, such as tanh+RP and DL+RT, our proposed two-stage transformation, RG+RP, maintains privacy of both normal and anomalous data, and delivers the lowest recovery rates. The privacy results demonstrate the robustness of RG+RP.

Furthermore, a novel LSTM-CNN framework for activity recognition is proposed. Experimental results demonstrate that it is substantially more accurate than the standalone LSTM or CNN model, and delivers a better performance than the DBN model. We believe this improvement is due to the time sequential encoding of activity features. Addtionally, when used with RG+RP, LSTM-CNN suffers an accuracy reduction of only 4.76% and 3.64% for the HAR and MH dataset respectively, while providing tremendous benefits in preserving privacy. The experimental results indicate that our proposed LSTM-CNN framework is useful for distinguishing activities that are performed by multiple subjects, which we believe is valuable for providing automated health monitoring and assistance in a smart environment. Consequently, LSTM-CNN combined with the

Table 5: Accuracy of the proposed LSTM-CNN under different hyperparameters. For CNN filters, (a,b) denotes b a-gram filters.

Model	HAR				MH			
	Raw		Perturbed		Raw		Perturbed	
Input Features	9	9	9	9	21	21	21	21
Hidden Units for LSTM	18/28	18/28	18/28	18/28	28/64	28/64	28/64	28/64
Layers for LSTM	1	2	1	2	1	2	1	2
CPU time per Mini-Batch	38s/40s	79s/85s	20s/25s	48s/53s	39s/43s	84s/90s	28s/31s	52s/59s
Filters for CNN	(30, 5), (40, 10), (50, 15), (60, 20)							
Accuracy	0.970/0.984	0.980/0.988	0.931/0.937	0.933/0.938	0.950/0.955	0.958/0.962	0.915/0.920	0.918/0.921

privacy-preserving scheme of RG+RP appears to be a good trade-off between accuracy and privacy.

The next steps from this work are plenty, including a formal analysis of how the parameters of the RG function and the pdf of the original data affect the recovery resistance of RG+RP; consideration of other kinds of attacks; and extensions of current work to other applications of the Internet of Things, such as analysing sensitive location data from multiple users on the server side to estimate the spatial distribution of pedestrians over a target area.

REFERENCES

[1] Gergely Ács and Claude Castelluccia. 2011. I have a dream!(differentially private smart metering). In *International Workshop on Information Hiding*. Springer, 118–132.

[2] Rakesh Agrawal and Ramakrishnan Srikant. 2000. Privacy-preserving data mining. In *ACM Sigmod Record*, Vol. 29. ACM, 439–450.

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. 2013. A Public Domain Dataset for Human Activity Recognition using Smartphones.. In *ESANN*.

[4] Oresti Banos, Rafael Garcia, Juan A Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio Rojas, Alejandro Saez, and Claudia Villalonga. 2014. mHealthDroid: a novel framework for agile development of mobile health applications. In *International Workshop on Ambient Assisted Living*. Springer, 91–98.

[5] Kanishka Bhaduri, Mark D Stefanski, and Ashok N Srivastava. 2011. Privacy-preserving outlier detection through random nonlinear data distortion. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 41, 1 (2011), 260–272.

[6] Ella Bingham and Heikki Mannila. 2001. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 245–250.

[7] Phil Blunsom, Edward Grefenstette, and Nal Kalchbrenner. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.

[8] Keke Chen and Ling Liu. 2005. Privacy preserving data classification with rotation perturbation. In *Data Mining, Fifth IEEE International Conference on*. IEEE, 4–pp.

[9] Ronald Cramer, Ivan Bjerre Damgøard, and Jesper Buus Nielsen. 2015. *Secure Multiparty Computation and Secret Sharing*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9781107337756 Cambridge Books Online.

[10] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment* 1, 2 (2008), 1542–1552.

[11] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 486–503.

[12] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*. Springer, 265–284.

[13] Sarah M Erfani, Yee Wei Law, Shanika Karunasekera, Christopher A Leckie, and Marimuthu Palaniswami. 2014. Privacy-preserving collaborative anomaly detection for participatory sensing. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 581–593.

[14] Raghu K Ganti, Nam Pham, Yu-En Tsai, and Tarek F Abdelzaher. 2008. PoolView: stream privacy for grassroots participatory sensing. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 281–294.

[15] Chris R Giannella, Kun Liu, and Hillol Kargupta. 2013. Breaching Euclidean distance-preserving data perturbation using few known inputs. *Data & Knowledge Engineering* 83 (2013), 93–110.

[16] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. 2013. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications* 18, 3 (2013), 326–340.

[17] Nils Y Hammerla, Shane Halloran, and Thomas Ploetz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. *arXiv preprint arXiv:1604.08880* (2016).

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[19] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V Pearson, Dietrich A Stephan, Stanley F Nelson, and David W Craig. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 4, 8 (2008), e1000167.

[20] Zhengli Huang, Wenliang Du, and Biao Chen. 2005. Deriving private information from randomized data. In *Proc. 2005 ACM SIGMOD international conference on Management of data*. ACM, 37–48.

[21] Yann LeCun and Yoshua Bengio. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361, 10 (1995), 1995.

[22] Kun Liu. 2007. *Multiplicative data perturbation for privacy preserving data mining*. Ph.D. Dissertation. University of Maryland, Baltimore County, Baltimore, MD.

[23] Kun Liu, Chris Giannella, and Hillol Kargupta. 2006. An attackerâĂŹs view of distance preserving maps for privacy preserving data mining. In *Knowledge Discovery in Databases: PKDD 2006*. Springer, 297–308.

[24] Kun Liu, Chris Giannella, and Hillol Kargupta. 2008. A survey of attack techniques on privacy-preserving data perturbation methods. In *Privacy-Preserving Data Mining*. Springer, 359–381.

[25] Kun Liu, Hillol Kargupta, and Jessica Ryan. 2006. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. *Knowledge and Data Engineering, IEEE Transactions on* 18, 1 (2006), 92–106.

[26] Lingjuan Lyu, Yee Wei Law, Sarah M Erfani, Christopher Leckie, and Marimuthu Palaniswami. 2016. An improved scheme for privacy-preserving collaborative anomaly detection. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 1–6.

[27] Olvi L Mangasarian and Edward W Wild. 2008. Privacy-Preserving Classification of Horizontally Partitioned Data via Random Kernels. In *Proc. International Conference on Data Mining (DMIN)*, Vol. 2. 473–479.

[28] Michael J McGrath and Cliodhna Ní Scanaill. 2013. *Sensor Technologies: Healthcare, Wellness and Environmental Applications*. Apress.

[29] Shibnath Mukherjee, Zhiyuan Chen, and Aryya Gangopadhyay. 2006. A privacy-preserving technique for Euclidean distance-based mining algorithms using Fourier-related transforms. *The VLDB JournalâĂŤThe International Journal on Very Large Data Bases* 15, 4 (2006), 293–315.

[30] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016), 115.

[31] Yingpeng Sang, Hong Shen, and Hui Tian. 2012. Effective reconstruction of data perturbed by random projections. *Computers, IEEE Transactions on* 61, 1 (2012), 101–117.

[32] Yingpeng Sang, Hong Shen, and Hui Tian. 2012. Effective reconstruction of data perturbed by random projections. *IEEE Trans. Comput.* 61, 1 (2012), 101–117.

[33] Elaine Shi, HTH Chan, Eleanor Rieffel, Richard Chow, and Dawn Song. 2011. Privacy-preserving aggregation of time-series data. In *Annual Network & Distributed System Security Symposium (NDSS)*. Internet Society.

[34] Fan Zhang, Li He, Wenbo He, and Xue Liu. 2012. Data perturbation with state-dependent noise for participatory sensing. In *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2246–2254.