2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering

Acies: A Privacy-Preserving System for Edge-based Classification

Wanli Xue[1,3], Yiran Shen*[,3], Chengwen Luo[2], Wen Hu[1,3] and Aruna Seneviratne[1,3]
[1]*University of New South Wales,* [2]*Shenzhen University,* [3]*Data61- CSIRO*
Email: {w.xue, wen.hu}@unsw.edu.au,{chengwen}@szu.edu.cn,{Yiran.Shen, Aruna.Seneviratne}@data61.csiro.au
* *Corresponding Author, Yiran Shen, Yiran.Shen@data61.csiro.au*

***Abstract*—In this paper, we propose Acies, a differential privacy based privacy-preserving classification system for edge computing to secure the classification models offloaded to edge devices. Acies supports popular classifiers such as Nearest Neighborhood, Support Vector Machine and Sparse Representation Classifier with a variety of feature selection methods. According to our evaluation on different datasets, classification models with Acies can be private and remain high utility. Acies achieves reliable privacy protection under reconstruction attacks with minimal impact on classification accuracy ($2\%-5\%$) only. Acies outperforms the naive input dataset perturbation methods by up to 30$\%$ higher classification accuracy when the privacy requirements of the applications is high (ϵ is less than 2).**

1. Introduction

Edge computing framework has been visioned and applied in many real-world applications like authenticating query [1]. Comparing with traditional centralized cloud, edge computing pushes application logic and the underlying data to the edge of the network with the aim of reducing latency, improving availability and scalability. Edge computing can ease the network reliance for real time applications, for instance, the authenticating process. Besides, the carefully tuned trade-off between computing and communication responsibilities between edge servers, trusted servers and untrusted services can enlarge the fault-tolerance, churn, elasticity and many others scale to millions of users, which is more suitable for current IoTs environment. However, this central-decentral shifting computing framework introduces new threats on data privacy therefore it forces to enhance the trust, privacy and autonomy requirements in computing applications.

The privacy risk of edge computing. Technology advances such as dedicated connection boxes deployed in most homes, high capacity mobile end-user devices and powerful wireless networks are always coupled closely with concerns on trust, privacy, and autonomy. Edge computing introduces controlling of applications, users' data, and services away from central nodes (the "core") to the other logical extreme (the "edge") of the Internet. Those out-of-range controls make significant contribution to the efficiency of edge computing, however, they also create new system security threats.

Privacy-preserving solutions. In this paper, we focus on the privacy leakage and the corresponding solutions for Machine Learning (ML). The privacy-preserving solutions designed for cloud computing cannot be directly migrated to edge computing scenarios due to their difference in network paradigm, i.e., the introduction of local (edge) servers. The exisiting solutions for privacy-preserving in cloud computing considered the case that the data owner outsources training of ML model which is computationally intensive to professional cloud service providers without revealing the privacy of data. However, considering the real deployment circumstances of edge computing services, the edge servers are not trusted. The threat from malicious users has not been well addressed as it aims for more complex scenarios.

In this paper, we propose Acies[1], based on differential privacy mechanism, to address generic ML privacy problems in the emerging edge computing scenarios.

Contribution. The contributions of this paper can be summarised as followed:

- We study the new problem of privacy-preserving classification under new edge computing scenario. A new threat model is proposed to analyze new issues brought by the new computational paradigm.
- Acies applies differential privacy protection in the feature selection stage and can be smoothly incorporated in most of the current classification services without any changes on classification work flow and system architecture.
- According to our extensive evaluation on multiple datasets of different classification tasks, Acies achieves reliable privacy protection against reconstruction attacks with only minimal impact on classification accuracy ($2\%-5\%$) which, is significant (up to 30%) lower than naive approaches by taking noise into account in the ML model training stage instead of adding the noise to the input data source directly.

2. Edge-based Classification

2.1. Characteristics and Constraints

Edge computing allows us to process data near the source and only send few results over the network to an

1. Acies is a Latin word origin as sharp edge and vision.

2324-9013/18/31.00 ©2018 IEEE
DOI 10.1109/TrustCom/BigDataSE.2018.00130

intermediate data processor, which can address unreliable latency problem in traditional cloud-based computing. There are couple of characteristics and constraints we need to consider when design an edge computing systems.

Dimension reduction and feature extraction. Training samples like features extracted from raw sensor recordings are used to train a robust classifier. IoTs devices (edge nodes) undertake some lightweight computing locally (we focus on classification tasks in this paper) to avoid latency caused by remote communication or network connection interruption. However, considering the fact that IoTs devices are resource-constrained, the complexity of trained ML models should be reduced through dimensionality reduction before being deployed on edge nodes. However, dimensionality reduction method should be carefully designed so that the network latency can be addressed without noticeable sacrifice on classification accuracy.

Kernel based classifiers. In this paper, we focus on the classifiers that are widely adopted in IoTs applications such as kNN, SVM and SRC [2]. All those classifiers can be categorized as Kernel Logistic Regression models (KLR) [3]. In KLR models, the kernel function retains (a tiny fraction of) the training data (termed as "import points") which may result in the leakage of privacy of the training data.

2.2. Threat Model

Different from traditional cloud computing threat model, we assume that the cloud server is secure while the edge (including edge servers and edge devices) is untrusted because the edge is located closer to the clients and difficult to be safeguarded physically like the Cloud. Furthermore, edge devices such as smartphones and smart home gateways are typically managed by clients without sophisticated cybersecurity knowledge to provide universal accessibilities, which makes them significantly easier to be compromised by the hackers.

We assume that users' personal devices (edge) are not compromised, otherwise the users are unlikely to adopt an extra system to provide a privacy preserved service. The classification model is trained on central server with tailored ML algorithms. In order to provide better system performance such as response time, the ML models are offloaded to edge servers and/or devices. A curious user or a malicious hacker may wish to obtain other user's private information. For example, Eve can access to the face authentication model placed in their building's edge server and may be interested in other residents' appearance.

2.3. Reconstruction Attack

To better illustrate the privacy issues of the edge computing based classification system we conduct some preliminary experiments by launching reconstruction attacks on those edge-based classifiers. We take the face recognition case as a visualized example within the whole paper.

Algorithm 1 Input Data Perturbation

1: **Input:** Normal input data/feature $X = \{x_1, x_2, x_3, ..., x_n\} \in \mathbb{R}^{n \times m}$, target privacy budget ϵ, extraction ratio k;
2: **Output:** Perturbed extracted/compressed feature set A ;
3: **Initialization:** Calculate sensitivity $s = \max X - \min X$;
4: For each $x_i \in X$
5: Sample noise vector lp from Laplace(s/ϵ)
6: $x_i' = x_i + lp$;
7: Computes the Singular Value Decomposition (SVD) of perturbed features transpose $(X')^T$: $(X')^T = U\Lambda V^T$,
8: Choose the first k column of U as the extraction matrix $R \in \mathbb{R}^{k \times n}$,
9: Get $A \in \mathbb{R}^{k \times m}$ by multiply the extraction matrix transpose with perturbed input $A = R^T \times X'$.

The attack for popular feature selection algorithms such as SVD, Eigenface and Fisherface is similar to those privacy attacks reported in the literature [4]. For the benefit of space, we omit the technical details.

Apart from face recognition authentication systems, an adversary can launch similar reconstruction attacks to other edge computing applications such as activity recognition. For instance, gait information extracted from WiFi signal like Channel State Information (CSI) can be used for human identification [5]. We will use these IoTs application datasets to evaluate the proposed algorithm later in Section 4.

3. Protections for Edge-based Classification

3.1. Input Data Perturbation

Inspired by differential privacy, a naive approach (as shown in Algorithm 1) can be applied to protect privacy by injecting random noises into classifiers' training dataset. The algorithm starts by initializing the sensitivity (Line 3 in Algorithm 1) then the noise vector is computed and combined with the chosen target privacy budget (Line 5). Selecting a reasonable global sensitivity itself is another research question which is beyond the scope of this paper. We adopted the naive approach from the original definition. Readers can refer to [6] for sensitivity methods like $\ell_2 - sensitivity$ for vector-valued functions. In face recognition case, the sensitivity is 255. According to the parallel composibility of differential privacy [7], the entire dataset X is differential private after processed by the algorithm. SVD is used as a feature extraction example in the algorithm, though other classifiers and feature extraction methods, such as Eigenface and Fisherface, are also applicable here. The output of this algorithm, i.e., A, then can be used for any ML classifiers as the training set.

Figure 1 presents examples of reconstruction attack results on different feature extraction methods (from top to bottom: Elgenface, Fisherface, SVD) under different rates of noise (ϵ=$\{8, 5, 2, \ln 2\}$) that are added in the training data. In differential privacy, smaller ϵ provides higher privacy guarantee, however, it may leads to lower usability, i.e., lower recognition accuracy. From the results shown in Figure 1

Algorithm 2 Feature Extraction Perturbation

1: **Input:** Normal input data/feature $X = \{x_1, x_2, x_3, ..., x_n\} \in \mathbb{R}^{n \times m}$, target privacy budget ϵ, extraction ratio k;
2: **Output:** Perturbed extracted/compressed feature set A ;
3: **Initialization:** Computes the Singular Value Decomposition (SVD) of perturbed features transpose $(X)^T$: $(X)^T = U\Lambda V^T$;
4: Choose the first k column of U as the extraction matrix $R \in \mathbb{R}^{k \times n}$,
5: Calculate sensitivity $s = \max R - \min R$,
6: **For** each $r_i \in R$
  7: Sample noise vector lp from Laplace(s/ϵ)
  8: $r_i' = r_i + lp$;
9: Get $A \in \mathbb{R}'^{k \times m}$ by multiply the extraction matrix transpose with perturbed input $A = (R')^T \times X$.

we can find that under small ϵ (such as ϵ=2 and ϵ=ln2), it is hard to recognize the identities of the face images after reconstruction attacks on the feature extraction methods of SVD and Eigenface. However, the added noises also have a huge impact on the recognition accuracy. In the literature, ϵ=ln2 is typically considered as providing acceptable level of privacy [8], but all ML algorithms lose their utility because the recognition accuracy drops to approximately 30% for kNN and SRC and 50% for SVM respectively as our evaluations in Section 4.4.

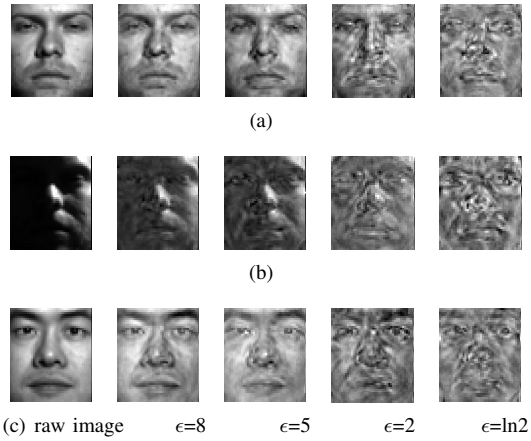


Figure 1: Face images reconstruction with different feature extraction methods after protection with Algorithm 1. (a)SVD (b)Eigenface (c)Fisherface.

3.2. Acies - Feature Extraction Perturbation

To achieve high privacy guarantee while preserving acceptable utility of the classification model in edge computing, we propose a new classification model perturbation method **Acies** (Algorithm 2). The fundamental idea behind is to perturb the tailored feature extraction methods, to control the information leakage from the training data indirectly.

Considering linear regression purpose, matrix R and X from Algorithm 2 can be also regarded as two independent linear models. Hence, the models of R and X then can be written as $\sigma = R(\lambda)$ and $b = X(a)$, where λ and a are different inputs to each model and σ and b are their corresponding outputs. Under the context of edge computing, we mostly use the combined model $(R(X(\cdot)))$ to handle the classification task, which refers to the feature selection process followed by those classifiers. Regardless of various classifiers, Acies takes linear transformation to generate the $R(X(\cdot))$, which equals to A in Algorithm 2.

We launch the reconstruction attacks to the features perturbed by Acies and present some preliminary results in Figure 2. The face images are the same to those used in Figure 1. The results show that the identities of the face images can be effectively protected when ϵ of R is equal to or below 8 which demonstrates significantly improved privacy preserving performance intuitively compared with the results shown in Figure 1, i.e., Acies added significant lower level of noise to provide reliable privacy protection compared with the naive approach.

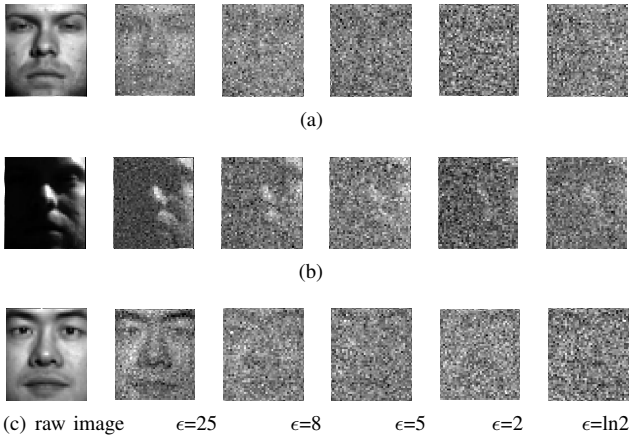


Figure 2: Face images reconstruction with different feature extraction methods after protection with Algorithm 2. (a)SVD (b)Eigenface (c)Fisherface.

4. Evaluation

4.1. Goals and Metrics

In this section, we will first illustrate how Acies affects ML models' utility by evaluating the recognition accuracy of Acies with a number of IoTs aplpication datasets, which include YaleB recognition (YaleB) [9] and WiFi Channel State information (CSI) [5].

We then evaluate the performance of Acies on privacy protection. We use mutual information [10] as our evaluation metric, which measures the distance between the original training instance and the estimation of the instance obtained from reconstruction attacks. The mutual information of two variables A and B can be computed using the probability distributions,

$$I(A,B) = \sum_{a,b} p_{AB}(a,b) log \frac{p_{AB}(a,b)}{p_A(a) \cdot p_B(b)} \qquad (1)$$

where $p_A(a)$ and $p_B(b)$ are marginal probability distribution and joint probability distribution $p_{AB}(a,b)$ are statistically independent if $p_{AB}(a,b) = p_A(a) \cdot p_B(b)$. When mutual information of two variables $I(A,B) = 0$, it implies that A and B are absolute independent. In the context of this paper, smaller $I(X,\hat{X})$ implies that X is better protected from reconstruction attacks. To illustrate Acies only introduce small system overhead to edge server/devices, we evaluate the time consumption of different classification models.

4.2. Recognition Accuracy of Classification System with Acies

YaleB We choose the first 32 face images from each class as training dataset and the following 10 as testing. 300-dimensional feature vectors are extracted for different classifiers. We compute the classification accuracy of the two privacy protection methods over the 38 classes. The averaging classification accuracy results are presented in Figure 3.

Compared with the input data perturbation approach shown on the left column of Figure 3, Acies (the right column) achieves significantly higher recognition rate with smaller ϵ. Acies has different impacts on different feature extraction methods, for example, the accuracy of SVM with Fishface drops to 7% (note that this data point is not shown in Figure 3(d)) while for other feature extraction methods, the change of ϵ has a less impact on accuracy.

CSI WiFi CSI data can be used to identify people based on the unique gait information. One data record is generated for a single person completing required activity. The raw CSI data is collected from 20 people, and each activity is repeated for 10 times. From the results (Figure 4) we can see that Acies has very little impact on classification accuracy for different values of privacy parameters ϵ.

In summary, Acies shows good performance on preserving the utility of different classification models with various feature extraction methods on different real-world datasets generally.

4.3. Privacy Analysis

In this section, we analyze the secrecy property of Acies under reconstruction attacks. We quantify how much information, in terms of mutual information, that an adversary can obtain from the classification models by launching the reconstruction attacks introduced in Section 2.3.

Figure 5 shows the normalized mutual information revealed via launching reconstruction attacks on the classification models perturbed by input data perturbation (Algorithm 1) and Acies (Algorithm 2) respectively for all datasets with 30% compression ratio. The results show that the normalized mutual information degrades with the decrease of ϵ (the amount of noise increases) for both methods as expected. However, the normalized mutual information of Acies drops significantly quicker than that of input perturbation algorithm which implies that the performance of Acies on privacy protection is significantly better. For example, when ϵ is 8, the mutual information of Acies is less than 0.2, while that of input data perturbation is approximately 0.3. We have also evaluated Acies and input data perturbation using other compression ratios (e.g., 10% and 50%). The results are similar to those in Figure 5, and the related plots are omitted for the benefit of space.

Table 1: Time consumption for classification process with different extraction ratio

Time Consumption						
Data Set/Time(s)	YaleB			CSI		
	5%	25%	100%	5%	25%	100%
kNN	0.006	0.01	0.02	0.05	0.13	0.71
SVM	0.9	1.1	1.2	0.32	0.36	1.87
SRC	0.3	0.67	1.58	0.3	0.9	2.3

4.4. Resource Consumption Analysis

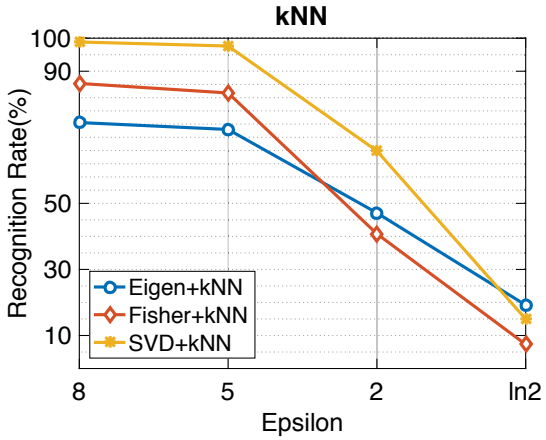
To evaluate the resources consumption of the classification systems with Acies, we measure the computation time and storage cost of different classification models with different datasets.

We conduct experiments on the same datasets as we used in the previous section, but for the benefit of space we select dimension reduction ratios of 5% and 25% as examples. As the results shown in Table 1, different dimension reduction ratios can affect the time consumption significantly for the same algorithms while time consumption is important for user experience (latency or system response time) issues. We take the classification using CSI as an example, the feature's original dimension is 10,800 (showed as 100%). When the feature dimension is compressed to 5%, the time consumption of classification model can be improved by 13, 6 and 8 times for kNN, SVM and SRC respectively. Similar results can also be observed from other datasets.

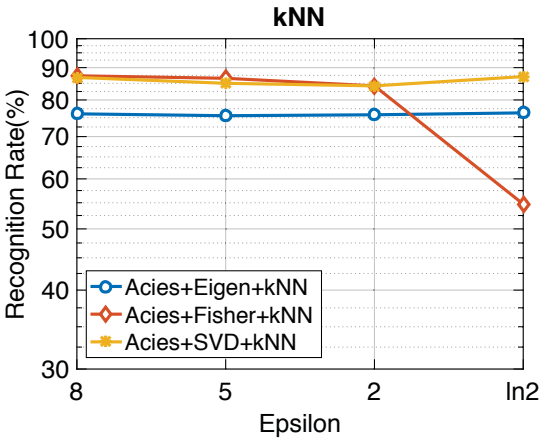
5. Related Work

In this paper, we perturb the classification models to preserve both the privacy and utility. Similar studies [11], [12], [13], [14], [15] have been conducted in the literature. However, the existing work is not applicable in edge computing environment as they focused on cloud computing scenarios which have significantly distinctive network architecture and application scenarios. The state-of-the-art privacy preserving mechanisms are either data-oriented (such as over partitioned dataset) [12], [13] or focus on specific ML model [14], [15], which require deliberate manipulation to make it as a fair comparison to the propose work.

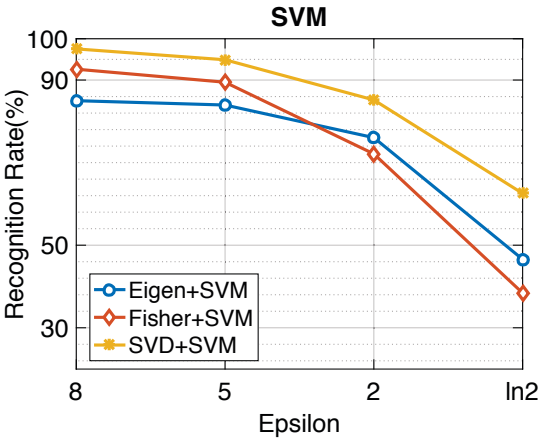
The first category of the related work was the studies on addressing the privacy leakage when releasing the classifiers. Lin and Chen [11] proposed to control the released features to prevent the privacy leakage. For conventional classification process with SVM classifier, all support vectors must be kept in the classifier, which might violate privacy. To protect the sensitive content in the classifier, a post-processing privacy-preserving SVM classifier schema was



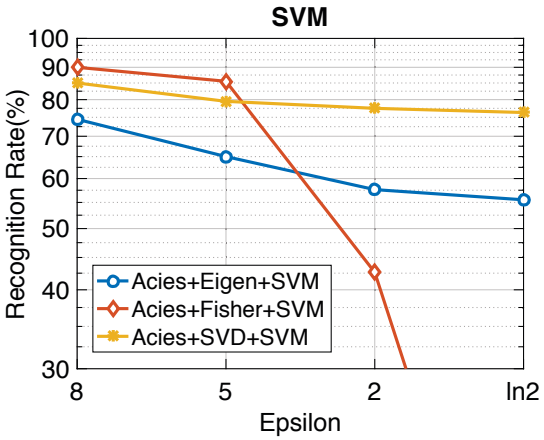
(a) NN with input data perturbation



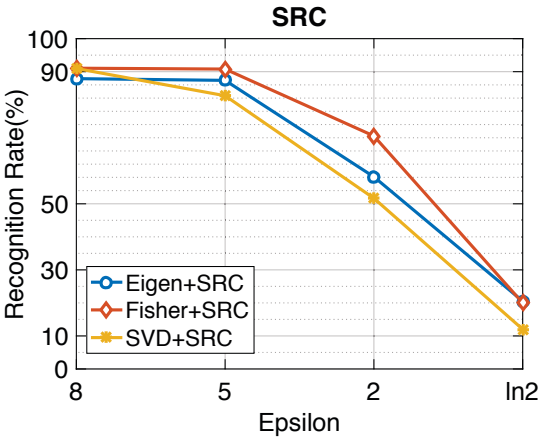
(b) NN with Acies



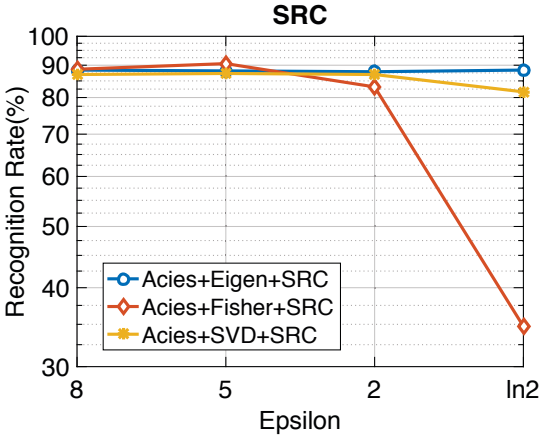
(c) SVM with input data perturbation (Note different Y-axis scale)



(d) SVM with Acies



(e) SRC with input data perturbation



(f) SRC with Acies

Figure 3: Recognition rates on extended Yale B database with input data perturbation (Algorithm 1) and Acies, for various feature transformations and classifiers.

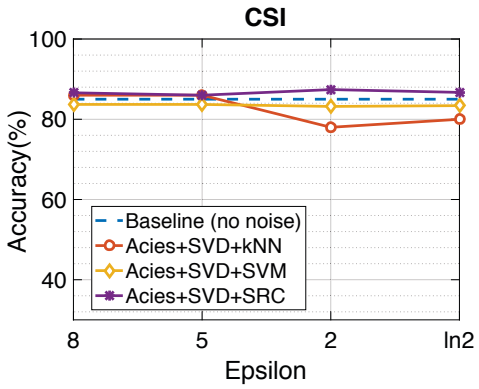


Figure 4: Classification accuracy on CSI dataset.

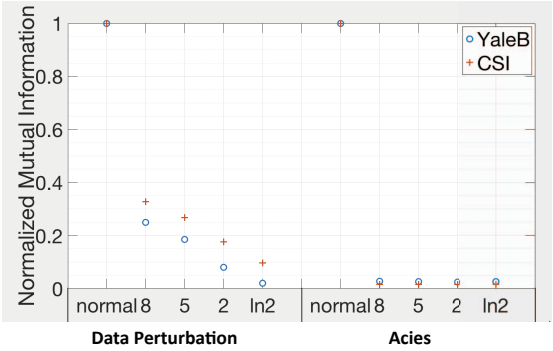


Figure 5: Effect of reconstruction attack on different datasets.

used to protect the sensitive content of support vectors in the classifiers. Many researches had also been proposed to protect the privacy leakage of other classifiers [12], [13], [14], [15] and they randomly perturbed the feature extraction methods in different classifiers. Those approaches required that all participants shared a common perturbation matrix to vertically/horizontally partition the private data/feature, where elements of the feature vector were spread among participants. Those methods are specific to classifiers and necessarily require fine-tuning from the model release party.

6. Conclusion

In this paper, we propose, Acies, a privacy-preserving system for classification on IoTs devices under edge computing environment. Instead of direct input data perturbation in the training set, which has huge impact on the classification accuracy, Acies is designed to perturb the feature extraction component to address the privacy issues when ML models are offloaded from the Cloud to edge nodes. By comparing the naive input data perturbation method on multiple datasets with different classification models, we can show that the proposed method, Acies achieves significantly better trade-off on privacy and utility preserving than naive approach: Acies provides trusted classification services with minimal impact on classification accuracy ($2\% - 5\%$).

References

[1] H. Pang and K.-L. Tan, "Authenticating query results in edge computing," in *Data Engineering, 2004. Proceedings. 20th International Conference on*. IEEE, 2004, pp. 560–571.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[3] J. Zhu and T. Hastie, "Kernel logistic regression and the import vector machine," *Journal of Computational and Graphical Statistics*, vol. 14, no. 1, pp. 185–205, 2005.

[4] S. P. Kasiviswanathan, M. Rudelson, and A. Smith, "The power of linear reconstruction attacks," in *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2013, pp. 1415–1433.

[5] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *Distributed Computing in Sensor Systems (DCOSS), 2016 International Conference on*. IEEE, 2016, pp. 75–82.

[6] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *Journal of Machine Learning Research*, vol. 12, no. Mar, pp. 1069–1109, 2011.

[7] F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. ACM, 2009, pp. 19–30.

[8] A. Friedman, S. Berkovsky, and M. A. Kaafar, "A differential privacy framework for matrix factorization recommender systems," *User Modeling and User-Adapted Interaction*, vol. 26, no. 5, pp. 425–458, 2016.

[9] Yale, "The extended yale face database b," 2001. [Online]. Available: http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html

[10] D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2001, pp. 247–255.

[11] K.-P. Lin and M.-S. Chen, "Releasing the svm classifier with privacy-preservation," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 899–904.

[12] O. L. Mangasarian, E. W. Wild, and G. M. Fung, "Privacy-preserving classification of vertically partitioned data via random kernels," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 2, no. 3, p. 12, 2008.

[13] H. Yu, J. Vaidya, and X. Jiang, "Privacy-preserving svm classification on vertically partitioned data." in *PAKDD*, vol. 3918. Springer, 2006, pp. 647–656.

[14] W. Xue, C. Luo, G. Lan, R. Rana, W. Hu, and A. Seneviratne, "Kryptein: a compressive-sensing-based encryption scheme for the internet of things," in *Information Processing in Sensor Networks (IPSN), 2017 16th ACM/IEEE International Conference on*. IEEE, 2017, pp. 169–180.

[15] J. Qiang, B. Yang, Q. Li, and L. Jing, "Privacy-preserving svm of horizontally partitioned data for linear classification," in *Image and Signal Processing (CISP), 2011 4th International Congress on*, vol. 5. IEEE, 2011, pp. 2771–2775.

[16] J. Vaidya and C. Clifton, "Privacy-preserving k-means clustering over vertically partitioned data," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 206–215.