PhD Forum Abstract: Towards Utility-Aware Privacy-Preserving Sensor Data Anonymization in Distributed IoT

Xin Yang
University of Alberta
Edmonton, AB, Canada
xin.yang@ualberta.ca

ABSTRACT

With the increased deployment of Internet of Things (IoT) systems, there come multiple privacy concerns regarding how the sensor data is stored and processed. Our research aims to protect user privacy by developing distributed data anonymization techniques that obscure sensitive information in the raw sensor data before it is used by third-party applications. Since users may have different privacy requirements, we propose to leverage meta-learning to allow our model to adapt to various privacy-utility trade-offs. We will assess the efficacy of the proposed techniques on real-world datasets and design defense mechanisms against various adversaries.

CCS CONCEPTS

• **Computer systems organization → Sensor networks**; • **Security and privacy → Privacy protections**.

ACM Reference Format:
Xin Yang. 2021. PhD Forum Abstract: Towards Utility-Aware Privacy-Preserving Sensor Data Anonymization in Distributed IoT. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3486611.3492389

1 INTRODUCTION

The proliferation of Internet of Things (IoT) and mobile sensing technologies has greatly improved user convenience and service availability. However, these advantages come at a cost; sensor readings can be easily exposed to an untrusted party or an adversary, thereby threatening user privacy. The adversary can infer the user's private attributes, such as gender, height, and weight, without their consent, for example by training a machine learning model on the motion sensor data that is originally gathered for step counting [2]. Data anonymization has proven to be an effective approach to obscure or remove sensitive information embedded in the sensor data so that the accuracy of inferring private attributes from the obscured data drops to the level of random guessing without sacrificing the data utility for desired tasks.

Existing anonymization techniques [3, 4] usually depend on a centralized, high-performance server to train its model on massive real-world datasets, requiring users to first disclose their sensitive raw data to this central server. This imposes privacy risks under a practical semi-trusted setting, where a trusted server can be compromised to infer sensitive private attributes. Various efforts have been made to preserve data privacy in large data systems using distributed frameworks like Hadoop [6]. Although they achieve k-anonymity and scale with the number of users, they are not designed to process streaming data from sensors. Recent research [5] utilizes Federated Learning (FL) to enable privacy-preserving model training in Trusted Execution Environments (TEE), but it requires dedicated hardware to deploy. While simply applying FL to machine learning-based anonymization models seems promising, there remain significant challenges when we consider practical applications. First, the dataset generated by a single user can be insufficient to properly train a data-driven anonymization model. Besides, users generate sensor data with feature distributions that can significantly differ from each other, causing biased local gradient estimations (i.e., the non-I.I.D. problem) and potentially decreasing the performance of the central model. Moreover, a universal central model can have very limited adaptability to fit the diversified privacy-utility trade-offs among a heterogeneous population.

My PhD aims to fill the gap in the literature and explore the design of a new privacy-preserving data anonymization technique that can be trained in a distributed fashion. Our insights are that the target task of building a centralized, highly adaptable data anonymization model can be learned from multiple sub-tasks residing at the participating users' local devices. Therefore, we aim to develop an algorithm that can aggregate a centralized anonymization model from the sub-tasks optimized at the user-end. Specifically, we plan to investigate the feasibility of leveraging Model-Agnostic Meta-Learning (MAML) [1] to avoid sharing raw sensor data and reduce the number of training samples required for each sub-task. The meta-learned anonymization model can be adapted to satisfy different users' privacy preferences, even for those who never contributed to the model training before. The contribution of my PhD research will be threefold:

- We propose a distributed privacy-preserving sensor data anonymization model leveraging meta-learning that can be trained on small datasets and adapted to fit the privacy-utility trade-off for a heterogeneous population.
- User studies and real-world experiments will be conducted to understand how users weigh privacy over utility. Metrics quantifying users' privacy-utility trade-offs will be developed to enable similarity-based user grouping that allows fine-grained model aggregation and personalized services.
- We plan to investigate the model security under model poisoning attacks when a subset of the participating users are malicious. Defense mechanisms based on adversarial networks will be developed.

BuildSys '21, November 17–18, 2021, Coimbra, Portugal
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9114-6/21/11.
https://doi.org/10.1145/3486611.3492389

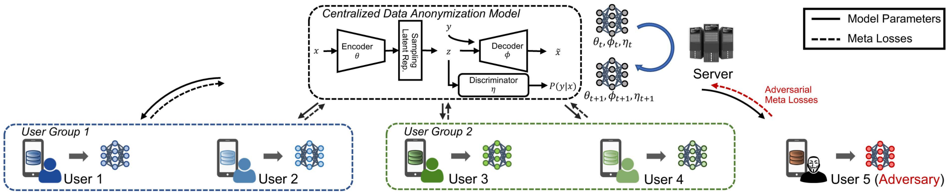

Figure 1: Overview of the proposed distributed data anonymization framework and attack models.

2 METHODOLOGY

2.1 Data Anonymization in Distributed Setting

We first aim to develop a new data anonymization model that can preserve user privacy during the model training on top of our prior work [3], which comprises an encoder, decoder, and discriminator as illustrated in Figure 1. Our model intends to address: 1) What information should be shared between users and the server to protect raw sensor data? 2) How to perform model aggregation to learn from the privacy preferences of heterogeneous users?

As the first step, we propose a distributed model training approach based on MAML, which allows a central model to learn from various sub-tasks residing in the participating users' local devices and quickly adapt to new tasks. The model training process contains two parts: distributed local model optimization and meta-learning-based model aggregation. Specifically, each participating user first fetches a copy of the central model and optimizes it locally using a small batch of data sampled from the user's local dataset. The optimization follows the adversarial training approach as described in [3]. Next, another batch of non-overlapping local data is sampled to compute the losses of the optimized discriminator and decoder (i.e., meta losses). In the model aggregation stage, the server collects the meta losses from a sufficient amount of participating users to update the central anonymization model through weighted averaging. Hence, the central model is optimized towards minimizing the distances between all participating users' local tasks, allowing any user's privacy preferences to be easily satisfied by performing adaptation on a small batch of local data.

2.2 Quantifying Privacy-Utility Trade-off

The second objective of our work aims to understand: 1) How different users weigh their privacy over the utility? 2) How to let service providers be aware of users' diversified privacy preferences and enable user grouping to offer next-level user experiences?

We plan to develop mobile apps to collect sensor datasets in real-world experiments. The collected datasets will be labeled based on the user feedback about their privacy-utility trade-offs assessed using subjective rating scales. Then, we intend to extract indicators representing the user's privacy preferences from the anonymized sensor data and the user's feedback. The indicators will be encrypted and shared with the central server during model training to inform the server of the user's preferences about the privacy-utility trade-off. Further, we aim to explore similarity metrics based on clustering and pairwise correlation to categorize users sharing similar privacy preferences as the user groups shown in Figure 1, thereby enabling many practical utility and privacy features, such as group-level model aggregation and inner-group differential privacy.

2.3 Security Threats and Defense Mechanisms

We further study the model security when adversaries pretend as legitimate users to send malicious updates to corrupt the central model (i.e., model poisoning attacks). We aim to study: 1) How to protect the integrity of the central anonymization model under the model poisoning attacks? 2) What is the maximum adversary tolerance of our model and how to improve it?

We plan to first model the attackers using generative adversarial networks (GAN) and let it arm wrestle with the obtained central anonymization model to generate adversarial updates that can reduce the anonymization performance of the central model. Then, malicious model updates will be collected and labeled to train a classifier that can identify and exclude malicious updates before the aggregation. We will evaluate how the number of adversaries could impact anonymization performance, and to enhance the adversary tolerance using adversarial examples and cryptographic algorithms.

3 CONCLUSION

While the market for IoT wireless sensors is expected to enjoy healthy growth over the next 5 years [7], data privacy and security related concerns could hinder this growth as users are becoming more aware of the privacy implications of installing IoT devices at home. For example, Amazon Alexa and Google Home have been accused of spying on users by recording their conversations to generate personalized advertisements. Unlike existing data anonymization techniques that typically ignore the privacy risks during model training, the proposed research aims to extend the protection of sensor data to its entire life cycle. Therefore, our work could strengthen users' trust in service providers and unleash the potential of data-driven IoT applications by allowing service providers to access anonymized real-world user data.

BIOGRAPHY

Xin Yang is a first-year Ph.D. student supervised by Dr. Omid Ardakanian in the CS Department at the University of Alberta. His Ph.D. dissertation will be submitted in 2025.

REFERENCES

[1] Chelsea Finn et al. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. PMLR, 1126–1135.

[2] Omid Hajihassani et al. 2020. Latent representation learning and manipulation for privacy-preserving sensor data analytics. In *2020 IEEE Second Workshop on Machine Learning on Edge in Sensor Systems (SenSys-ML)*. IEEE, 7–12.

[3] Omid Hajihassnai et al. 2021. ObscureNet: Learning Attribute-invariant Latent Representation for Anonymizing Sensor Data. In *Proc. International Conference on Internet-of-Things Design and Implementation*. 40–52.

[4] Ang Li et al. 2021. DeepObfuscator: Obfuscating Intermediate Representations with Privacy-Preserving Adversarial Learning on Smartphones. In *Proc. International Conference on Internet-of-Things Design and Implementation*. 28–39.

[5] Fan Mo et al. 2021. PPFL: privacy-preserving federated learning with trusted execution environments. In *Proc. 19th Annual International Conference on Mobile Systems, Applications, and Services*. 94–108.

[6] J Jesu Vedha Nayahi and V Kavitha. 2017. Privacy and utility preserving data clustering for data anonymization and distribution on Hadoop. *Future Generation Computer Systems* 74 (2017), 393–408.

[7] Research and Markets. 2021 [Online]. IoT Wireless Sensor Market by Type, Technology, Solutions and Applications in Industry Verticals 2021 - 2026. https://www.researchandmarkets.com/reports/5390546/