CAOE | Center for Accelerating Operational Efficiency

A Department of Homeland Security Center of Excellence

Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise

June 21, 2021

Arizona State University

Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise
June 21, 2022 • Washington, DC

Privacy-enhancing technologies (PETs) under development promise the ability to control the sharing and use of sensitive information while minimizing the risk of unauthorized use. These technologies have been under development by researchers for nearly four decades but have been slow to migrate from the research lab into operational use.

This document collects a set of white papers that were selected to be presented at the Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise. The workshop received 21 submissions, of which 12 were ultimately selected for presentation.

Organizers:

Ross Maciejewski – Arizona State University
Simson Garfinkel – Department of Homeland Security

Program Committee

Jalal Mapar - DHS
Rene Peralta - NIST
Albert Cheu - Georgetown
Rachel Cummings - Columbia
Rafail Ostrovsky - UCLA
Meaghan Catalano - DHS
Robert Beverly - NSF
Tatiana Ringenberg - Indiana University
James Michael - Naval Postgraduate School
Michele Steinmetz - DHS
Jonathan Katz - George Mason University

Disclaimer

This event is hosted by the Center for Accelerating Operational Efficiency, a Department of Homeland Security Center of Excellence in collaboration with the Department of Homeland Security Privacy Office.

This material is supported by the U.S. Department of Homeland Security under Grant Award Number, 17STQAC00001-03-03.

Disclaimer. The views and conclusions contained in this document should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security

For more information, contact CAOE at CAOE@ASU.edu or visit https://caoe.asu.edu/.

Table of Contents

Secure Federated Learning

Jose-Luis Ambite, Srivatsan Ravi, Greg Ver Steeg
Information Sciences Institute, University of Southern California

There are situations where data relevant to a machine learning problem are distributed across multiple locations that cannot share the data due to regulatory, competitiveness, security, or privacy reasons. Federated Learning (FL) is a promising approach to learn a joint neural network model over all the available data across silos without transferring data to a centralized location. However, federated learning is challenging. In many cases, the sites participating in a federation have different data distributions, computational, and communication capabilities. In these heterogeneous environments existing approaches exhibit poor performance, including slow model convergence, high energy cost, or high communication requirements. Moreover, securing the data and the neural model from attacks that would disclose sensitive information is critical.

We have developed a ***Secure Federated Learning architecture that learns efficiently in heterogenous environments and provides strong security guarantees***. This architecture was recently developed under funding from the DARPA Artificial Intelligence Exploration Cooperative Secure Learning program [1,2,3,4,7]. Figure 1 shows a sketch of the architecture. Each site (learner) trains on its local dataset for some time, then sends its neural model parameters (e.g., gradients, weights) to a Federation Controller, which aggregates the local models from multiple sites into a community model. This community model is sent back to the sites for continued training and the process repeats. The architecture is general; it supports a wide variety of federated training policies, neural network architectures (e.g., convolutional, recurrent, transformers), and data types (e.g., images, text, structured, multimodal).

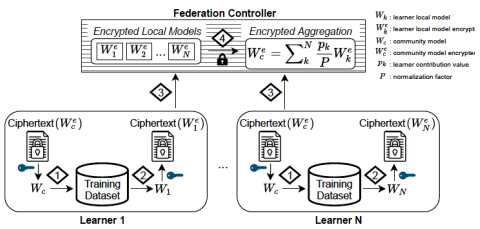


Fig 1. Secure Federated Learning [2]

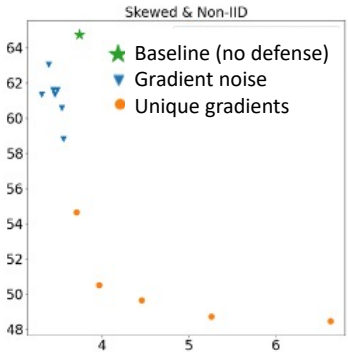


Fig 2. Information-theoretic defenses against membership attacks. Our non-unique gradients method outperforms differential privacy. x-axis: Task Error (lower is better), y-axis: Membership attack success % (lower is better).

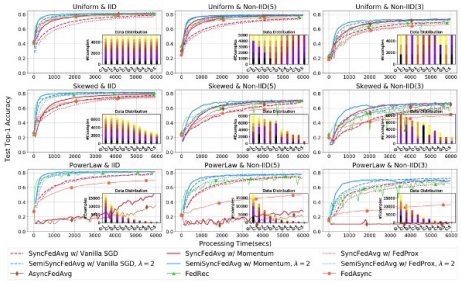


Fig 3. Efficient Federated Learning in heterogenous environments [1]

Our architecture is secure and keeps data at all sites private. Data never leaves a site. Model parameters are encrypted before transmission. Transmission of model parameters between a site and the federation controller is through secure channels. The community model is computed under fully-homomorphic encryption using the CKKS construction [8], so even if the controller was compromised the community model cannot be attacked. In [2] we show that learning with or without encryption achieves the same learning accuracy, with a modest increase in computation time (~15%). We are also investigating a Pallier scheme that splits encryption into offline and online phases, which has significantly less overhead during (online) federated training.

Unfortunately, current deep learning models are highly susceptible to inversion and membership attacks [3], which can recover information about training data given only access to the model. When a site receives the community model, it decrypts it for further training, thus a *curious* site could perform an attack. A membership attack consists of discovering whether some sample data was used to train a model. If the task pursued by the federation is sensitive, this is unacceptable. For example, if the federation task is person or object identification in video or images, a site can probe with its own images to reveal if any of them was present in other sites in the federation, which could be a security breach.

We have been pursuing several methods to ***prevent information leakage from neural models***:

- *Information-theoretic privacy guarantees*. We have developed a method that preferentially add noise that obscures unique information contributed by individual samples. Figure 2 shows experimental results where our method performs better than differential privacy. While differential privacy is the gold standard for protecting data, in practical deep neural networks training scenarios enforcing differential privacy often destroys task performance [3]. Additionally, theoretical results show that information-theoretic methods can improve task performance by bounding the generalization error for deep neural networks [9].
- *Invariant representation learning*. We learn representations of data that are invariant to sensitive attributes. In our experiments in neuroimaging, we find that each fMRI scanner has a unique fingerprint and "erasing" this information improves task performance, while also improving privacy [10]. In security applications, we do not expect threats to depend on the type of surveillance equipment used for detection. Therefore, we may purposely learn representations that are invariant to the equipment both to obscure sensitive information about the data source and to improve the generalizability of the model to different data sources.
- *Sparsification*: Sparsifying/pruning models (both in federated and non-federated settings) is another way to reduce over-fitting that also has privacy benefits. We plan to investigate the privacy guarantees of our work on progressive sparsification for federated training [7]. This training policy has the additional benefit of reducing communication costs.

We have developed methods for efficient federated training [1,4,7]. Figure 3 shows how our training policies (blue lines, top-left) converge faster than alternatives in heterogeneous data distributions (see plot insets) and computational environments. Due to their fast convergence, our policies are 3 to 9 times more energy efficient in reaching a given accuracy [1].

We propose to apply our Secure Federated Learning methods to enable distributed analysis of sensitive data across agencies without sharing data and ensuring that no information from a site can leak into any other site. Our architecture is containerized to facilitate deployment. We have demonstrated the benefits of our approach in biomedical domains where the protection of patient data is paramount. Since our architecture is general, it can be applied to other sensitive domains. Possible DHS applications include: 1) analysis of distributed image or video assets that must remain private (e.g., object, entity, or activity detection in surveillance or satellite images, fake-video detection); 2) cross-silo NLP over distributed private documents (e.g., entity or event extraction in intelligence reports, detection of spear-phishing/spam across private emails servers, private sentiment analysis); 3) multisite analysis of security events to detect intrusions (here in addition to privacy concerns, the large amount of local data makes it impractical to transmit data to a centralized site for analysis); and 4) private distributed multimodal data analysis (e.g., joint image, text, sensor data analysis).

[1] Stripelis, Ambite, Thompson. **Semi-Synchronous Federated Learning for Energy-Efficient Training and Accelerated Convergence in Cross-Silo Settings**. *ACM TIST*. In press, doi:10.1145/3524885. 2022.

[2] Stripelis et al. **Secure neuroimaging analysis using federated learning with homomorphic encryption.** *Int'l Symp Medical Information Processing and Analysis (SIPAIM)*, 2021.

[3] Gupta, Stripelis, Lam, Thompson, Ambite, ver Steeg. **Membership inference attacks on deep regression models for neuroimaging**. *Medical Imaging with Deep Learning (MIDL)*. 2021.

[4] Stripelis Ambite, Lam, Thompson. **Scaling neuroscience research using federated learning**. In *IEEE International Symposium on Biomedical Imaging (ISBI),* Nice, France, 2021.

[5] Harutyunyan et al. **Improving generalization by controlling label-noise information in neural network weights**. In *International Conference on Machine Learning (ICML),* 2020.

[6] Jaiswal, Brekelmans, Moyer, Ver Steeg, AbdAlmageed, Natarajan. **Discovery and separation of features for invariant representation learning**. arXiv:1912.00646, 2019.

[7] Stripelis, Gupta, ver Steeg, Ambite. **Federated Progressive Sparsification (Purge, Merge, Tune)+**. arXiv:2204.12430. 2022.

[8] Cheon, J. H., Kim, A., Kim, M., and Song, Y., **Homomorphic encryption for arithmetic of approximate numbers**, in *Advances in Cryptology, ASIACRYPT 2017*, Takagi, T. and Peyrin, T.,eds., 409-437. 2017.

[9] Harutyunyan, Raginsky, Ver Steeg, Galstyan. **Information-theoretic generalization bounds for black-box learning algorithms**. *Advances in Neural Information Processing Systems 34 (NeurIPS).* 2021.

[10] Moyer, Ver Steeg, Tax, Thompson. **Scanner invariant representations for diffusion MRI harmonization**. *Magnetic Resonance Medicine* 84: 2174– 2189. 2020

Privacy-preserving Graph Analytics: Secure Generation and Federated Learning

Dongqi Fu[†], Jingrui He[†], Hanghang Tong[†], and Ross Maciejewski[§]
[†]University of Illinois at Urbana-Champaign, [§]Arizona State University
{dongqif2, jingrui, htong}@illinois.edu, rmacieje@asu.edu

ABSTRACT

Directly motivated by security-related applications from the Homeland Security Enterprise, we focus on the privacy-preserving analysis of graph data, which provides the crucial capacity to represent rich attributes and relationships. In particular, we discuss two directions, namely privacy-preserving graph generation and federated graph learning, which can jointly enable the collaboration among multiple parties each possessing private graph data. For each direction, we identify both 'quick wins' and 'hard problems'. Towards the end, we demonstrate a user interface that can facilitate model explanation, interpretation, and visualization. We believe that the techniques developed in these directions will significantly enhance the capabilities of the Homeland Security Enterprise to tackle and mitigate the various security risks.

1 INTRODUCTION

Nowadays, the Homeland Security Enterprise is facing unprecedented challenges in multiple critical areas, such as identifying and preventing targeted violence and mass attacks, building resilient critical infrastructure, countering human trafficking, etc. Addressing these challenges requires collaborative efforts from all levels of government, the private and nonprofit sectors, and individual citizens. In particular, effective and efficient data collection, sharing, analysis, and sense-making are at the core of many decision making processes in these areas. Due to the distributed, sensitive and/or private nature of the large volume of involved data (e.g., personal identifiable information, images and video from surveillance cameras or body camera), it is thus of great importance to make use of the data while avoiding the sharing and use of sensitive information. In this paper, we focus on graph data, or network data, due to their rich representation capabilities to encode the multi-modality time-evolving attributes for entities (e.g., individuals, locations) as node attributes in the graph, and to encode the various types of relationships among entities via edges.

The main goal of this paper is to focus on privacy-preserving analysis of graphs. In particular, we aim to explore key research questions and privacy-enhancing technologies that target the following two areas: (A1) privacy-preserving techniques for creating, maintaining and linking anonymous identities and profiles; and (A2) multiparty and homomorphic computation to allow for analysis of datasets held by multiple parties without the need to physically combine datasets. The techniques developed for these areas together will enable the collaboration among multiple parties each possessing private graph data.

More specifically, we consider the following two directions for the two areas respectively. For (A1), a promising direction is to generate synthetic graphs in a privacy-preserving manner, such that the generated graphs can be shared with collaborators without revealing sensitive information on either the node level (node-DP) or the edge level (edge-DP). For (A2), a promising direction is to perform federated learning among multiple parties, or clients, such that the central server can build robust, effective and efficient predictive models while preserving the privacy of individual clients. Next, we elaborate on the two directions, including both 'quick wins' and 'hard problems', followed by additional discussions regarding the development of the user interface that could enable subject matter experts to make effective use of the developed techniques.

2 PRIVACY-PRESERVING GENERATION

Graphs represent complex relational information between entities, such that modeling the graph generation process and then generating many more meaningful graphs could contribute to various applications [3]. However, mimicking the observed graph as much as possible will induce a privacy risk after the generation [11]. For example, a node's identity is highly likely to be exposed in the generated social network if its connections are mostly preserved, which means a degree-based node attacker will easily detect a vulnerability in the generated graph with some background knowledge.

2.1 Quick Wins

For privacy-preserving static graph generation, current solutions can be roughly classified into two types that can be readily applied. First, the anonymization is directly performed on the observed topology to generate new graph data, such as randomizing the adjacency [14], injecting the connection uncertainty [9], or permutating the connection distribution under the edge-level differential privacy (edge-DP) [10]. Second, following the synergy of deep learning and differential privacy [1], deep generative models for graphs are recently proposed under privacy constraints. To be specific, in [13], the privacy scheme is added to the gradient descent phase of the generation learning process.

2.2 Hard Problems

Most, if not all, of privacy-preserving graph generation methods consider the observed graphs as static. However, in the complex real-world scenarios, the graphs are usually evolving over time [4], which brings critical challenges to the current privacy-preserving static graph generation process. To the best of our knowledge, how to generate privacy-preserving temporal graphs largely remains open. For example, (1) unlike the abundant research on static graphs, what kind of time-aware information is sensitive and should be hidden in the generated graph to protect entities' privacy is not clear; (2) even if the sensitive information is determined, the time-aware protection mechanism is not yet available; (3) once the protection mechanism is designed, it can be challenging to maintain the generation utility at the same time with privacy constraints.

3 FEDERATED LEARNING WITH GRAPHS

According to [6], "Federated learning (FL) is a machine learning setting where many clients (e.g. mobile devices or whole organizations) collaboratively train a model under the orchestration of a central server (e.g. service provider), while keeping the training data decentralized." This problem setting is particularly important for privacy-preserving analysis of private graph data that might reside on multiple clients such as the servers for various organizations. In other words, it enables centralized decision making using all available graph data, while avoiding physically combining datasets.

In the past few years, federated learning has been extensively studied, focusing on research problems such as model robustness to adversarial attacks, and data distribution among clients. Despite some existing work using graph data (e.g., [5]), robust federated learning with graph data largely remains under-explored.

3.1 Quick Wins

In general, federated learning systems can be vulnerable to attacks and failures, e.g., Byzantine attacks can lead to the convergence to an unsatisfactory model or even divergence [2]. A common defending mechanism is to replace the gradients averaging with robust estimation of the center [2]. These methods have proven Byzantine-robustness when data from different clients are independent and identically distributed (IID). On one hand, these methods can be directly adapted to model graph data by using, e.g., graph neural networks [12]. On the other hand, the IID assumption regarding client data can fail miserably due to the special properties of graph data (e.g., homophily) as compared to other types of data.

To address this problem, one 'quick win' would be to observe the performance of existing Byzantine-robust methods on graph data, in order to study the impact of IID violation to the performance of the centralized model. Based on the results from this study, another 'quick win' would be to develop robust federated learning methods tailored for graph data that violate the IID assumption. For example, in our previous work, we studied a special type of non-IIDness among client data, i.e., label skewness, and proposed a two-stage algorithm to estimate the true parameters of the centralized model in the presence of Byzantine clients. Based on this work, one can further study other types of non-IIDness among client data that may fit various types of graph data, and design the robust algorithms.

3.2 Hard Problems

For privacy-preserving analysis of graph data, one major benefit of federated learning is that it could avoid physically combining private data from individual clients. On the other hand, some recent studies on federated learning require the availability of clean and non-sensitive data on the server (e.g., to enable knowledge distillation in the presence of heterogeneous models among clients [7]), which can potentially be shared with clients without violating the privacy constraints. However, such data can be difficult to obtain, especially in security related applications (e.g., the identification of targeted violence). Despite some recent work focusing on data-free knowledge distillation for federated learning [15], the proposed solutions cannot be readily applied on graph data due to the significant difference in generating graph data vs. non-graph data. To solve this 'hard problem', the graph generative models (as discussed in the previous section) for both static and dynamic graphs can potentially be integrated into the federated learning pipeline, to enable the data-free robust federated learning for graph data.

4 USER INTERFACE

With the development of novel privacy-enhancing technologies, one critical aspect is model explanation, interpretation, and visualization facing various audience. This calls for effective and efficient user interfaces that serve three major purposes: (1) to solicit feedback from subject matter experts for model improvement; (2) to empower subject matter experts with deep insights regarding the model and the data; (3) to aid decision making in various security related applications. The following figure exemplifies such an interface that we developed for explaining transfer learning [8].

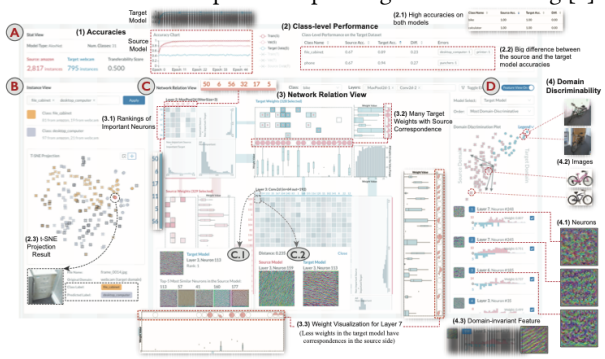


Figure 1: Visual Analytics for Transfer Learning Interface

5 CONCLUSION

In this paper, we discussed privacy-preserving analysis of graph data in two directions, namely secure generation and federated learning. We believe that research efforts dedicated to these directions can lead to significantly enhanced capabilities of the Homeland Security Enterprise for countering the various security risks.

REFERENCES

[1] M. Abadi, A. Chu, I. J. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. 2016. Deep Learning with Differential Privacy. In *CCS 2016*.

[2] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer. 2017. Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. In *NeurIPS 2017*.

[3] A. Bonifati, I. Holubová, A. Prat-Pérez, and S. Sakr. 2020. Graph Generators: State of the Art and Open Challenges. *ACM Comput. Surv.* (2020).

[4] D. Fu and J. He. 2021. SDG: A Simplified and Dynamic Graph Neural Network. In *SIGIR 2021*.

[5] C. He, K. Balasubramanian, E. Ceyani, Y. Rong, P. Zhao, J. Huang, M. Annavaram, and S. Avestimehr. 2021. FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks. *CoRR* (2021).

[6] P. Kairouz and et al. 2019. Advances and Open Problems in Federated Learning.

[7] T. Lin, L. Kong, S. U. Stich, and M. Jaggi. 2020. Ensemble Distillation for Robust Model Fusion in Federated Learning. In *NeurIPS 2020*.

[8] Y. Ma, A. Fan, J. He, A. Reddy Nelakurthi, and R. Maciejewski. 2021. A Visual Analytics Framework for Explaining and Diagnosing Transfer Learning Processes. *IEEE Trans. Vis. Comput. Graph.* (2021).

[9] H. H. Nguyen, A. Imine, and M. Rusinowitch. 2015. Anonymizing Social Graphs via Uncertainty Semantics. In *CCS 2015*.

[10] Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren. 2017. Generating Synthetic Decentralized Social Graphs with Local Differential Privacy. In *CCS 2017*.

[11] X. Wu, X. Ying, K. Liu, and L. Chen. 2010. A Survey of Privacy-Preservation of Graphs and Social Networks. In *Managing and Mining Graph Data*.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. 2019. A Comprehensive Survey on Graph Neural Networks. *CoRR* (2019).

[13] C. Yang, H. Wang, K. Zhang, L. Chen, and L. Sun. 2021. Secure Deep Graph Generation with Link Differential Privacy. In *IJCAI 2021*.

[14] X. Ying and X. Wu. 2008. Randomizing Social Networks: a Spectrum Preserving Approach. In *SDM 2008*.

[15] Z. Zhu, J. Hong, and J. Zhou. 2021. Data-Free Knowledge Distillation for Heterogeneous Federated Learning. In *ICML 2021*.

Twin Finder: Discovering and Assessing the Vulnerability to AI-Generated Twin Identities

Mohamed Hussein, *Research Lead, USC Information Sciences Institute*

Introduction

Privacy enhancing technologies (PETs) enable sharing of personal information while preserving privacy. PETs consider personal information that is obtained by the consents of the owners. However, there is a recent trend in AI that makes it possible for personal identifiable information to be released and widely shared without the consent of the owner. Modern artificial intelligence (AI) methods have the ability to create pictures that match the quality of natural images, including synthesizing photo-realistic face images of people who do not exist in real life. However, there is no guarantee that such people do not exist. What if one of these AI-generated faces is an identical twin of a real person?! Such *fake twins* can be used in ways their real counterparts never consented for, which can, intentionally or unintentionally, cause harm.

Meanwhile, while modern AI models have been able to match or exceed human performance in multiple tasks, such as face recognition and playing chess, they have also been found to be vulnerable to making completely incorrect decisions upon imperceptible perturbations to their inputs. For example, a face recognition model can be fooled into identifying an image of person A as person B if specific maliciously-crafted imperceptible perturbations are applied to person A's image. This phenomenon underscores another type of AI-generated twins. These twins are generated to "look" to an AI model as their real counterparts while they do not look the same way to humans. Such *adversarial twins* are easier to generate and are potentially more harmful than the fake twins.

Existing research focuses on generating more naturally looking synthetic or adversarial face images without addressing the problem of discovering and assessing the vulnerability of individuals to the threats posed by these images. At USC Information Sciences Institute, we are developing a prototype system (Twin Finder) that focuses on helping communities and law enforcement agencies discover and assess the vulnerability of individuals to fake and adversarial twins.

The Twin Finder System

We are developing the Twin Finder web service, which employs deep learning methods to investigate existing AI models by searching for AI-generated twins. The Twin Finder framework and design are illustrated in Figure 1. The front-end of the service will allow users to upload one or more query face images. After processing the user's input, the system will generate a vulnerability report. The front-end will securely communicate with a backend server. The Twin Finder server will query a set of face

generation models for the closest matching *fake twins* by solving an inverse optimization problem. It will also inspect a set of face matching models to find the closest *adversarial twins*. Adversarial twins reported by the system will fall in two categories: (1) imperceptibly perturbed versions of the query image that reliably fail matching with the query image, (2) imperceptibly perturbed versions of other face images that can reliably be matched with the query image. For the latter type of adversarial twins, the system will use a gallery composed of publicly available large face image datasets. The gallery will be indexed to be efficiently searched for the closest matches before creating the adversarial perturbations. The service will allow for uploading multiple faces for the same person, which will help produce more reliable results. All images in the produced vulnerability report will be accompanied with their matching scores with the closest query image. From these matching scores and the number of found matches, an overall vulnerability score will be estimated.

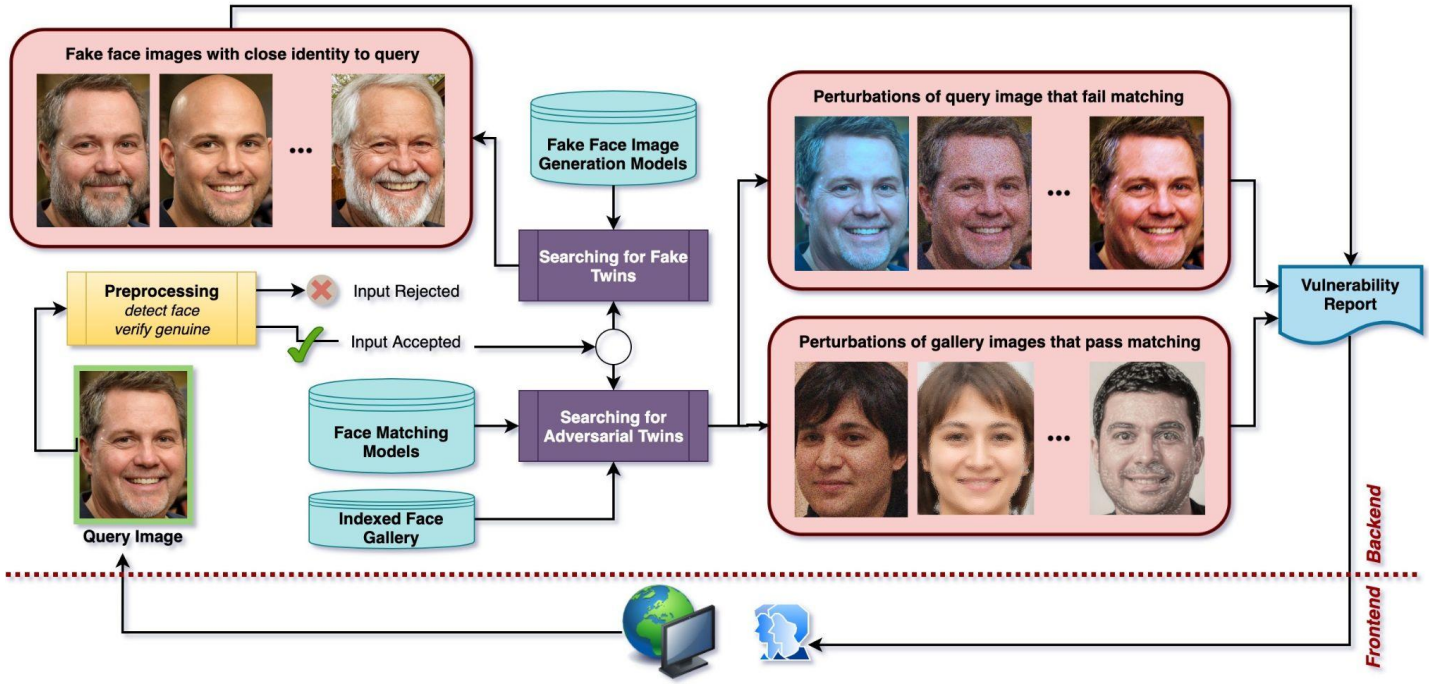


Figure 1. Twin Finder's Schematic Diagram

Evaluation: The frontend interface will request user's feedback on the quality of the produced results in terms of how realistically-looking the produced images are and how similar or dissimilar they are to the query images. The user's feedback will be a critical part of our system's evaluation. Cases in which the user's feedback significantly deviates from the system's estimates will be investigated for possible system enhancement.

About the Author: The author is a co-PI on USU ISI's efforts under IARPA's BRIAR program (on biometrics) and DARPA's GARD program (on adversarial robustness).

Final Note: Please, do not include in the conference material if the idea is not selected for discussion.

Privacy-Preserving Video Surveillance System over Cloud

Vishesh Kumar Tanwar, Sanjay Madria, and Sajal K. Das
Department of Computer Science
Missouri University of Science and Technology Rolla, Missouri

1. Motivation

Privacy-Preserving Video Surveillance system (PPVSS) is an end-to-end framework for secure computations over the captured data to accomplish the dedicated utility task such as activity recognition and object(s) tracking while protecting the private information of the unintentionally captured individuals. For instance, PPVSS installed at an airport can recognize an occurrence of suspicious activity while protecting travelers' faces. Numerous privacy-enhancing technologies (PETs) in video surveillance have developed in the last few years. Still, they are limited to protecting only a few private regions like human faces and vehicle number plates. However, an individual's sensitive information also contains work and home locations, gender, race, clothes, religious affiliations, health issues, etc. In addition, the storage and computational advancements in cloud computing enhance the existing surveillance frameworks to store and manage the captured data on the cloud server in the plain domain (PD).

The cloud servers can utilize the transmitted data without users' consent which increases the concerns of users' private information leakage. The concerns of secure transmission and storage of the surveillance data on cloud servers have been addressed by the classical encryption schemes. However, these encryption schemes cannot be used for secure computations, like training a deep convolutional neural network (D-CNN), for the encrypted data over the cloud server. The bottleneck of D-CNNs is leveraging the pixels' inter-correlations of the input video frames to construct feature vectors followed by a classifier. Contrasting, the traditional image encryption schemes break the pixels' inter-correlations and incorporate a noise-like structure. In this paper, we propose to design and implement PETs to develop end-to-end PPVSSs (shown in Fig. 1), protecting individuals' private information without compromising the surveillance robustness for the next 3 to 5 years.

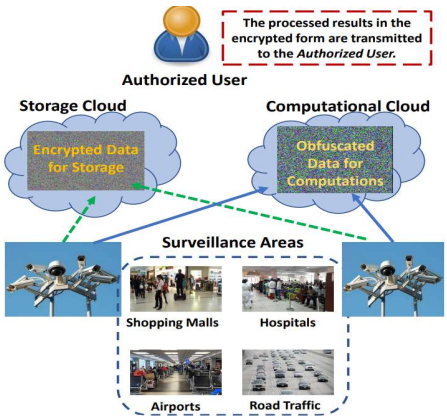


Figure 1: An end-to-end pictorial scenario of PPVSS.

1.1 Research Challenges

Researchers [1] have presented a few task-specific PETs for image processing, protecting the dedicated privacy attributes to ease the computations over the cloud servers. The leakage of sensitive private attributes, other than faces, can easily be observed in Fig. 2. A few secure computation frameworks incorporate homomorphic encryption (HE) schemes with multi-party computations (MPC). However, these schemes lack with three major drawbacks - *storage*, *computation* and *communication latency* overheads, making them impractical for real-time PPVSS. A little research is done on incorporating sensor-based security rather than secure computations on videos. Moreover, researchers utilized differential privacy (DP) and federated learning (FL) for training the D-CNNs in the local systems to protect the users' data privacy. However, these schemes deal with two significant issues: (1) DP creates a trade-off between the data security and the model's accuracy, and (2) FL is vulnerable to data reconstruction attack by the cloud server [2,3].

Figure 2: Existing PETs protect only dedicated region of interest like faces.

1.2 Missing Research Gaps

Summarizing the research gaps in PETs for video surveillance are (1) lack of defined privacy and security protocols/techniques, (2) Detection of privacy attributes in the encrypted domain (ED) with applications to object/pedestrian tracking, and (3) secure computations algorithms to develop efficient real-time surveillance models in the centralized and decentralized cloud computing.

2. Initial Methodology and Results

Our initial work developed an end-to-end PET to train an image recognition model for users' obfuscated data (using our designed approach) over a centralized server. Specifically, we implemented an

image obfuscation scheme aiming to preserve the image's local features and obfuscate the global spatial information, as shown in Fig. 3 Contrasting existing methods, we utilized a bottom-top approach incorporating non-invertible noise using the fractional-order chaotic system to develop a block-based image obfuscation scheme. A significant difference between our scheme in Fig. 3 and existing schemes (Ref. Fig. 2) can be observed in protecting complete image information. We have experimented with our obfuscation algorithm on two datasets - CIFAR10 and CIFAR100 and achieved recognition accuracy of 80% and 74%, respectively, using ResNet50 as a recognition model. From a security perspective, the obfuscated image obtained using our scheme cannot be reconstructed due to the utilization of DP-noise and thresholding. Moreover, our scheme prevents visual layer attacks, protection from impersonation, man-in-the-middle, and physical or stolen attack. Our scheme is secured under standard security attacks protecting the image information. Also, the system parameters need not be kept secret to achieve privacy guarantees.

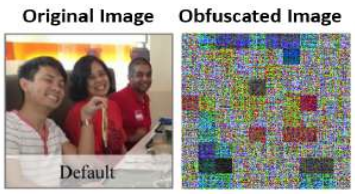


Figure 3: Obfuscated image obtained by our proposed approach.

3. Research Objectives: We propose to design the following three PETs to develop the end-to-end real-time PPVSSs, in a manner shown in Fig. 4:

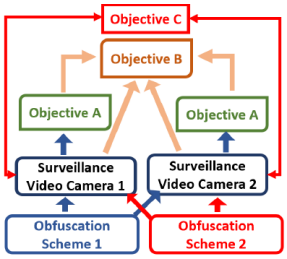


Figure 4: Block-diagram depicting the relations of objectives A, B and C for two surveillance cameras.

A. Privacy-Preserving Human Activity Recognition (HAR): We are interested in developing an end-to-end PET for HAR by encrypting the spatial features while preserving the temporal information only. Existing HAR PETs in video surveillance have two phases - (i) data capturing and (ii) data encryption, then transmission to the server for processing. The time window between the two phases is highly vulnerable to attack. Therefore, we will design HAR PETs to capture encrypted surveillance data, which can be directly transmitted to the server for processing. Moreover, our PETs will incorporate with the existing surveillance systems to recognize the suspicious activities in the ED.

B. Secure FL for Activity Recognition: Due to the diversified locations of video surveillance systems, the existing systems are not robust for real-time suspicious activity recognition. Moreover, state-of-the-art recognition systems utilize D-CNNs, which face domain adaptation problems. Therefore, it is required to develop an FL-based framework while protecting individuals' private information from the central server (as mentioned above), which we will accomplish as our second research problem. Extending (A), we will incorporate HE and DP with the camera lens performing computations locally on the installed camera in the encrypted form, followed by the model's weights-aggregation on the central server.

C. Tracking Road Activities in the Obfuscated Domain: We will develop a cloud-based PET to track road activities such as a vehicle(s) and pedestrian(a) detection and tracking while protecting private information, like vehicle number plate and location, passenger/pedestrian faces, etc. Previously, we have developed PET for traffic monitoring in VANETS [4]; however, PET for object detection is challenging because of maintaining object(s) pixel's inter-correlation during secure computations for same object reconstruction whenever required. Like, after detecting accidents in the ED, the vehicle number plate and passenger(s) face are required in the PD to provide medical help or make fines. Existing detection PETs are limited to the traditional pixel-permutation-based approach which are not computationally and communicationally robust for high-resolution surveillance videos. We will utilize chaotic systems and DP to develop a real-time on-site irreversible anonymization function to address these concerns. Further, we will utilize PET of (B) to improve robustness in diversified locations. Our PPVSS model will detect the moving object(s) in the obfuscated data over the cloud server, raise the alarm whenever a violation occurs, and provide the object information to the authorized person.

References

[1.] Bentafat, Elmahdi, M. Mazhar Rathore, and Spiridon Bakiras. "Towards real-time privacy-preserving video surveillance." Computer Communications, Elsevier, Vol. 180, pp. 97-108, 2021.
[2.] Geiping, J., Bauermeister, H., Dröge, H. and Moeller, M., 2020. Inverting gradients-how easy is it to break privacy in federated learning?. Advances in Neural Information Processing Systems, 33, pp.16937-16947.
[3.] Kong, X., Gao, H., Shen, G., Duan, G. and Das, S.K., 2021. Fedvcp: A federated-learning-based cooperative positioning scheme for social internet of vehicles. IEEE Transactions on Computational Social Systems.
[4.] Roy, A. and Madria, S., 2020, December. Secure and privacy-preserving traffic monitoring in VANETs. In 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS) (pp. 567-575). IEEE.

Self-supervised Deep Learning for Privacy-preserving Video Analytics

Chen Chen, Mubarak Shah, Ishan Rajendrakumar Dave
Center for Research in Computer Vision
University of Central Florida

Introduction

Video-analytics-as-a-service enables a wide range of real-world applications, e.g., video surveillance, smart shopping systems like Amazon Go, elderly person monitoring systems. A key concern in such services is the privacy of the videos being analyzed, as analyzing such information-rich video data may reveal personal information like an individual's daily routine, home location, gender, race, clothes, etc. Therefore, there is a pressing need for solutions to privacy-preserving video analysis.

A simple-yet-effective solution for privacy preservation is to use very low-resolution videos via down-sampling. However, the performance of the utility application such as action recognition would suffer greatly due to significant information loss. Another set of methods use pretrained object-detectors to detect the privacy regions and then remove or modify the detected regions using synthesis [1] or blurring [2]. However, the detection-based approaches require the bounding-box level annotations for the privacy attributes, which are time-consuming and expensive to obtain. The most recent work [3] proposed to learn an anonymization function through an adversarial training framework to remove the privacy features from videos. However, this method also requires the annotation of privacy attributes (e.g., skin color, face, gender, etc.) from videos to perform the supervised training.

Proposed Approach

In light of the limitations of the existing privacy-preserving video analysis methods, we, for the first time, proposed a novel training framework which removes privacy information from input video in a self-supervised manner without requiring privacy labels. This work [4] was accepted in CVPR 2022. Our method has been successfully applied to privacy-preserving video action recognition – one of the most popular applications in video analytics.

Instead of just focusing on the cues based on the privacy annotations, our goal is twofold: 1) learning an anonymization function that can remove all spatial cues in all frames without significantly degrading action recognition performance; and 2) learning the anonymization function without any privacy annotations. To achieve this goal, we introduce a novel self-supervised learning framework for privacy preserving action recognition. The key idea of our proposed framework is to learn an anonymization function such that it deteriorates the privacy attributes **without requiring any privacy labels in the training stage**, and maintains the performance of action recognition task. The proposed framework mainly consists of three components as shown in Fig. 1: (1) Anonymization function (f_A); (2) Self-supervised privacy removal branch; and (3) Action recognition or utility branch.

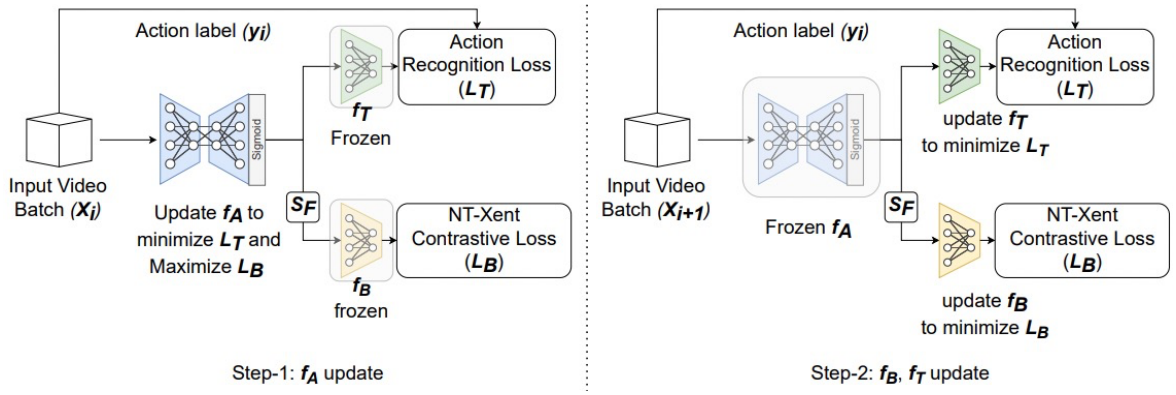


Fig. 1. A schematic of our framework for privacy preserving video action recognition. Please refer to our paper [4] for more details.

Specifically, the anonymization function is a learnable transformation function, which transforms the video in such a way that the transformed information can be useful to learn action classification on any target model, but not useful to learn any privacy identification model. A novel self-supervised learning method is developed based on the contrastive learning framework [5] to maximize the agreement between two frames of a same video and maximize the disagreement between frames of different videos. By maximizing the contrastive loss, it changes the input in such a way that it decreases agreement between frames of the same video. Intuitively, we know that frames of the same video share a lot of semantic information, and minimizing the agreement between them in the feature representation space results in destroying (i.e., unlearning) most of the semantic information of the input video. Since this unlearned generic semantic information contained privacy attributes related to human, scene, and objects; we end up removing private information in the input. In the meantime, the action recognition or utility branch is used to keep the useful information for maintaining the action recognition performance.

Our proposed method achieved the state-of-the-art results in terms of action-privacy trade-off on several benchmark datasets as reported in [4]. Fig. 2 presents a visual example of the anonymized video using our self-supervised privacy preservation method. It clearly shows that the private information has been successfully removed. However, a standard action recognition method can still recognize the action "ApplyLipstick" accurately based on the anonymized video.

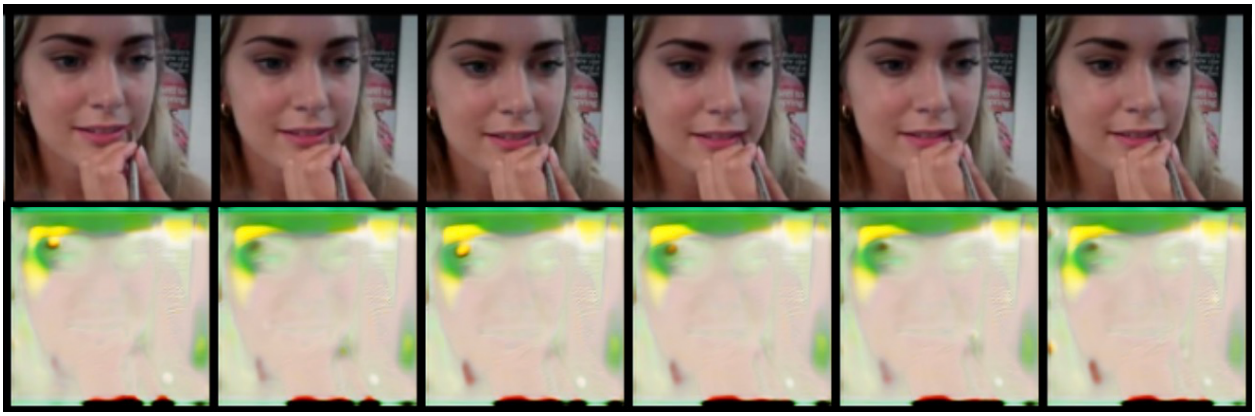


Fig. 2. First row: original video; second row: anonymized video using our self-supervised privacy preservation method. The standard action recognition model can still correctly recognize the action class as "ApplyLipstick" based on the anonymized video.

Future Research Agenda

In our future work, we plan to extend our self-supervised privacy preserving method to other video analysis tasks such as temporal action localization, action detection, and video object segmentation. Our proposed method can be incorporated with the encryption approaches to further improve privacy-preserving sharing and analysis of images and video. Moreover, our proposed method can be applied in the federated learning paradigm, where the anonymization transformation in our approach can protect the data on local clients and combat data leakage.

References

[1] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. Learning to anonymize faces for privacy preserving action detection. In Proceedings of the european conference on computer vision (ECCV), pages 620–636, 2018.

[2] Zhixiang Zhang, Thomas Cilloni, Charles Walter, and Charles Fleming. Multi-scale, class-generic, privacy-preserving video. Electronics, 10(10):1172, 2021.

[3] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[4] Dave, Ishan Rajendrakumar, Chen Chen, and Mubarak Shah. "SPAct: Self-supervised Privacy Preservation for Action Recognition." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. Preprint is available at: https://arxiv.org/pdf/2203.15205.pdf

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In ICML, 2020.

Privacy-preserved capturing and processing of images and videos

Rakibul Hasan

Arizona State University (rakibul.hasan@asu.edu)

Problem space

Images and videos contain rich information about the environment and surrounding people, whether or not those people were captured intentionally (i.e., subjects in those photos or videos) or incidentally (i.e., bystanders or passersby) [7, 6, 10, 9]. While obfuscating privacy-sensitive visual content is possible by, e.g., image filters or encryption [5, 8], the primary hurdle to overcome is to automatically understand what information should be considered as 'private' in a given context. In the case of surveillance or wearable cameras used by the members of law enforcement agencies, this issue is further complicated by the fact that some data that was filtered out may be required later to aid the investigation. Thus, one obvious first step towards privacy-preserving recording and analyses of image and video data in this context seems to be ensuring the privacy of bystanders, who are unrelated to the incidents. While conceptually straightforward, separating 'bystanders' from 'subjects' is extremely challenging as these constructs depend on the context and are difficult to categorize automatically based only on visual data. In an experiment, we asked human participants to categorize people in images as bystanders or subjects; we found that different participants labeled the same person differently for almost half of the images (N=5000) [1].

Our proposed solution and its potential impact

We leveraged cognitive science, machine learning, and computer vision to provide a fully automated mechanism to detect bystanders in images. We developed a classifier that mimics human reasoning in detecting bystanders using only image data so that bystanders' privacy can be protected by, e.g., obfuscating their faces or bodies with image filters. We created a separate test dataset to evaluate this model. Our model demonstrated 91.2% accuracy for images where all annotators agreed on a single class label, and 78.6% accuracy for images in which the roles of subjects and bystanders were ambiguous even to the human annotators (only 67% agreement). This fully automated and generic classifier can be readily deployed to cloud servers to process images and videos after uploading them, and on mobile and wearable devices (for in-device processing through model compression). Additionally, this tool can be used to sanitize already stored images and videos. We also studied image and video obfuscation methods to identify which filtering techniques are effective to obscure different sensitive content: such as paper documents, electronic displays, and the surroundings [2, 3, 4]; all of such information is frequently captured by body-worn cameras [7, 6, 9].

Scientific basis of our methods.

Through a user study, we gained insight into how humans conceptualize 'subjects' and 'bystanders' in images, based on which we identified and validated a set of intuitive features that 'make a person

a bystander or a subject.' These features were inferred from raw image data using several deep learning models and then used to train the final classifier. Additionally, we experimented with other models with different architectures and feature sets (e.g., fine-tuning large deep learning models), and all of them performed worse than the one described above. We hypothesize that our approach to identifying high-level features based on human intuition improves the signal-to-noise ratio of the input data which results in better classification accuracy by the model. Finally, the decisions made by our model can be easily explained as they are based on a few intuitive features; thus, our model ensures transparency and reliability which are much-sought characteristics of automated analyses in any domain, but particularly in policing and surveillance.

References

[1] Rakibul Hasan, David Crandall, and Mario Fritz Apu Kapadia. Automatically Detecting Bystanders in Photos to Reduce Privacy Risks. In *2020 IEEE Symposium on Security and Privacy (SP)*, Los Alamitos, CA, USA, 5 2020. IEEE Computer Society.

[2] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 47:1–47:13, New York, NY, USA, 2018. ACM.

[3] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. Can privacy be satisfying? On improving viewer satisfaction for privacy-enhanced photos using aesthetic transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, volume 14, page 25. ACM, 2019.

[4] E T Hassan, R Hasan, P Shaffer, D Crandall, and A Kapadia. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1333–1342, 7 2017.

[5] Jianping He, Bin Liu, Deguang Kong, Xuan Bao, Na Wang, Hongxia Jin, and George Kesidis. PuPPIeS: Transformation-Supported Personalized Privacy Preserving Partial Image Sharing. In *IEEE International Conference on Dependable Systems and Networks*, Atlanta, Georgia USA, 2014. IEEE Computer Society.

[6] Bryce C. Newell. Body cameras help monitor police but can invade people's privacy. https://theconversation.com/body-cameras-help-monitor-police-but-can-invade-peoples-privacy-160846/, 2021. Accessed: 2022-03-29.

[7] Bryce Clayton Newell. *Police visibility: Privacy, surveillance, and the false promise of body-worn cameras*. Univ of California Press, 2021.

[8] Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. P3: Toward Privacy-preserving Photo Sharing. In *USENIX Conference on Networked Systems Design and Implementation*, nsdi'13, pages 515–528, Berkeley, CA, USA, 2013. USENIX Association.

[9] Ethan Thomas. The privacy case for body cameras: The need for a privacy-centric approach to body camera policymaking. *Colum. JL & Soc. Probs.*, 50:191, 2016.

[10] Natalie Todak, Lindsay Leban, and Lois James. Citizen attitudes towards the public release of police body-worn camera video footage. *Police Practice and Research*, 22(7):1760–1776, 2021.

Privacy Enhancing Technologies Ready for the Homeland Security Enterprise

David W. Archer, PhD - Galois, Inc.

The Foundations for Evidence-Based Policymaking Act of 2018 (H.R.4174) mandates data sharing to promote informed decision-making, but current approaches to sharing sensitive data either put the data at risk or critically limit the utility of that sharing. Privacy-Enhancing Technologies (PETs) can enable high-utility data sharing while eliminating risks to data confidentiality, yet this opportunity has not yet been broadly proven because most PETs remain naïve with regard to operational security, bewilderingly difficult to use, fragile to operate, or thousands of times too slow. Here, we describe PET implementations positioned for "quick wins": either ready for transition today or that can be fully ready within a 2-year horizon. We are ready to demonstrate these tools on real use cases at DHS and submit them to critical review of their suitability for purpose. The PETs we describe are *Private Set Intersection with Secure Computation* (PSI), and *Zero Knowledge Proofs* (ZKPs). Galois' extensive experience in these and other PETs covers 10 years and many successful DHS, DARPA, and IARPA projects.

PSI

Applicability. Applies to DHS challenges such as privacy-preserving

- Construction or application of watch lists based on multiple data sources
- Linking anonymous identities and profiles
- Linking / analysis of data held by multiple parties without physically combining datasets

Readiness[1]. TRL 7 (Successfully demonstrated in an operational environment). The software is part of the Galois, Inc. open source PET library. We are actively seeking mission partners.

Overview. Our pilot project for the US Dept. of Education fully and accurately replaced a statistical application that requires sharing of sensitive data among activities within that Department[2]. A non-programmer intern at the Dept. used our tools to reproduce a portion of the annual 2015–16 National Postsecondary Student Aid Study (NPSAS:16) report, showing statistics on federal Title IV aid received by undergraduates in 31 categories for that academic year. The pilot was conducted in the normal environment of the Dept. of Education network.

The pilot used the same source data used by the original application. Data comes from two activities within the Department: the National Postsecondary Student Aid Study group (NPSAS) at the National Center for Education Statistics (NCES) and the National Student Loan Data System (NSLDS). Today, preparing the NPSAS reports requires that NSLDS must share sensitive student financial information with NPSAS, and NPSAS must share students' college study program information and social security numbers with NSLDS. To avoid those disclosures while successfully and efficiently providing the same statistics, our prototype uses PSI and

[1] According to the DHS Commercialization Office Technology Product Realization Chart, found at https://www.dhs.gov/xlibrary/assets/product_realization_chart.pdf

[2] *https://mccourt.georgetown.edu/news/a-federal-government-privacy-preserving-technology-demonstration/*

multi-party computation to link data and compute statistics without revealing sensitive data of either party to each other (or to anyone else).

Specifically, we produced accurate results for average Federal Pell Grants, Subsidized Federal Direct Loans, Unsubsidized Federal Direct Loans, and all Federal Direct Loans across institution type, attendance pattern, and income level. The full-year workload ran in a few hours, with computation and network burdens well within practical limits, and demonstrated that users without significant programming experience or cryptographic expertise can use PETs to protect data privacy in a production environment inside the Department's network.

In other settings for the US Government, we use slight variants of this PSI capability at TRL 6 to test whether any of a number of parties using network resources has a claim (for example, an operational dependency) on an IP address; to de-conflict kinetic operation use of physical locations during operational exercises; and to identify similarity between software executables. A related capability achievable with this technology is the search for certain "triggers" in text or audio while keeping the original media private, and revealing (perhaps only under court order) only the portion of the sensitive dataset that matched one or more triggers.

Applicability to specific DHS use cases seems natural, with limited additional work. PSI enables for example watch lists to be compared by matching on common identifiers and revealing only matching entries; watch lists to be compared against passenger manifests, revealing only passengers that "hit" on the list; and aggregation of profile information by linking on common identifiers, and revealing/aggregating only information on matching profiles. PSI as we provide it also enables linking with general analytics over multiple datasets without revealing those datasets or physically combining them. Additional privacy measures to protect data from compromise in the analytic results of these computations, such as *differential privacy*, can also be included in the PSI analytic functionality.

ZKPs

Applicability: Applies to DHS challenges where data is sensitive, but choices based on that data must be proven to meet agreed-on criteria without revealing the data to a verifying party (such as an auditor, or external watchdog). Examples include:

- Proof that a person meets the criteria to appear on a watch list without revealing PII
- Proof that a FISA court warrant was carried out correctly without revealing gathered data

Readiness: TRL 5 (All technology components are integrated, Proof of Concept conducted). Our tools are on the path to open source release. We are actively seeking mission partners.

Overview: As part of the DARPA SIEVE program, we have built and demonstrated the world's fastest, most scalable system for complex zero knowledge proofs. Successful, integrated proofs-of-concept have demonstrated proofs of vulnerabilities in software (such as the TLS Heartbleed vulnerability); proofs of memory safety of programs; and proofs that cryptographic material is prepared according to correct protocol. Proofs such as those outlined above are very likely to fit easily within the current capability of these tools.

VaultDB: Facilitating Secure Analytics over Multiple Private Data Sources

Xiao Wang
wangxiao@northwestern.edu

Jennie Rogers
jennie@northwestern.edu

Abstract

Data is changing practically every aspect of how the government does business. As agencies collect data for a myriad of endeavors, from research to operations, this is creating a treasure trove of insights waiting to be surfaced. To date, most of this data is siloed at its site of origin owing to privacy or security concerns. Here, data custodians have a legitimate need to maintain exclusive access to their sensitive data. On the other hand, these organizations would benefit from collaborating with one another when their interests and expertise overlap. In this text, we describe VaultDB, a system that queries the union of two or more private databases using cryptographic protocols such that the only information revealed is that which can be deduced from the query's answer. We also describe our results from deploying this technology within healthcare institutions for clinical research over electronic health records.

VaultDB is a data analytics platform for jointly querying the records in two or more datastores while keeping the secret inputs from each engine private. Like conventional data federations, it makes multiple autonomous database systems available for querying with a unified SQL interface. The client queries the union of these databases as if all of their data were stored in a single engine. VaultDB translates its queries into secure computation protocols that the data providers evaluate jointly for a given query. Hence, the private data never leaves the data owner's site, and the only information revealed from a query is that which can be deduced from its output. This output is only accessible to the individual who submitted the query.

VaultDB's high assurance of security is backed by cryptographically secure multi-party computation (MPC). MPC is an important subfield of cryptography studying protocols to enable two or more parties to jointly compute a function over their private inputs without divulging them to others. Here, the computing parties simulate a completely trustworthy, incorruptible third party by passing encrypted messages amongst themselves. VaultDB uses the Efficient Multi-Party (EMP) Toolkit [7] as its back end. This open-source project was started in 2016 and is being widely used and actively maintained. The toolkit was developed to address two goals: 1) provide an easy way for researchers to build efficient cryptographic prototypes; 2) provide a user-friendly interface for non-cryptographic developers to write privacy-preserving applications.

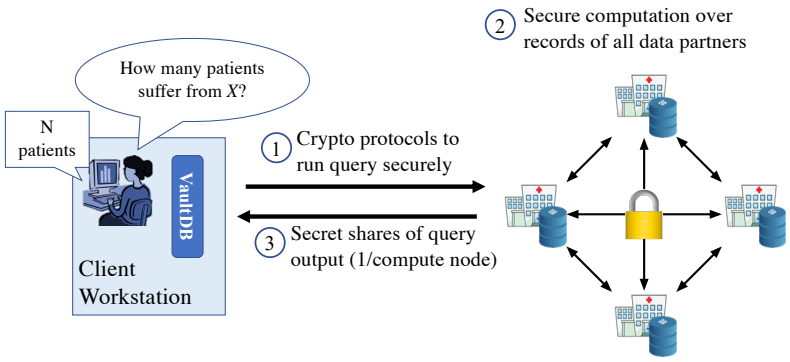


Figure 1: VaultDB workflow. The analyst runs VaultDB on her machine and inputs Q. The system reveals A to her alone and the data owners learn nothing about the secret inputs of their peers.

Figure 1 demonstrates its workflow with an example query that counts the distinct employees that joined all organizations in 2022. VaultDB connects seamlessly with existing database systems for SQL querying. We released an open-source prototype of this, SMCQL [2]. In previous work, we verified the utility of SQL over secure computation in a clinical research workload [1]. Since then, we have improved efficiency by integrating differential privacy [3] and approximate query processing via input sampling [4].

VaultDB is well-positioned to support many government operations including finding collaborators, searching for relevant private records, and exchanging intelligence from two or more sources on an entity of interest. It can easily be queried such that it only divulges information if two or more agencies are independently collecting information on the same entity.

Real-World Applications. In addition, our team has substantial expertise in bringing this technology to real-world settings. We recently developed and deployed a HIPAA-compliant version of VaultDB [6] on the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) [5], multi-institutional clinical research network. We conducted a study on hypertension rates over 600,000 patient records in the Chicago metropolitan area by deploying on-site at three health systems within this network show its efficiency and scalability for distributed clinical research analyses without moving patient records from their site of origin.

References

[1] Johes Bater, Gregory Elliott, Craig Eggen, Satyender Goel, Abel Kho, and Jennie Rogers. SMCQL: secure querying for federated databases. *Proceedings of the VLDB Endowment*, 10(6):673–684, 2017.

[2] Johes Bater, Gregory Elliott, Craig Eggen, Satyender Goel, Abel Kho, and Jennie Rogers. SMCQL: Secure querying for federated databases. `https://github.com/smcql/smcql`, 2017.

[3] Johes Bater, Xi He, William Ehrich, Ashwin Machanavajjhala, and Jennie Rogers. Shrinkwrap: Efficient SQL Query Processing in Differentially Private Data Federations. *Proceedings of the VLDB Endowment*, 12(3):307–320, 2018.

[4] Johes Bater, Yongjoo Park, Xi He, Xiao Wang, and Jennie Rogers. SAQE: Practical Privacy-preserving Approximate Query Processing for Data Federations. *Proceedings of the VLDB Endowment*, 13(12):2691–2705, 2020.

[5] Abel N Kho, Denise M Hynes, Satyender Goel, Anthony E Solomonides, Ron Price, Bala Hota, Shannon A Sims, Neil Bahroos, Francisco Angulo, William E Trick, et al. Capricorn: Chicago area patient-centered outcomes research network. *Journal of the American Medical Informatics Association*, 21(4):607–611, 2014.

[6] Jennie Rogers, Elizabeth Adetoro, Johes Bater, Talia Canter, Dong Fu, et al. VaultDB: A Real-World Pilot of Secure Multi-Party Computation within a Clinical Research Network. *arXiv preprint arXiv:2203.00146*, 2022.

[7] Xiao Wang, Alex J. Malozemoff, and Jonathan Katz. EMP-toolkit: Efficient MultiParty computation toolkit. `https://github.com/emp-toolkit`, 2016.

Rapid Prototyping of Secure Multi-Party Computation Applications

Emily Shen, R. Nicholas Cunningham, J. Parker Diamond, Noah Luther, David A. Wilson, David Bigelow
MIT Lincoln Laboratory

For many applications, multiple parties want to collaborate to gain insights from their collective data but don't want to share their own data due to security and privacy concerns. Current approaches to collaboration require the parties to trust each other or a third party with their sensitive data. These approaches are often undesirable, impractical, or impossible. Ideally, all parties could obtain the benefits of data sharing without actually sharing their data.

Secure multi-party computation (MPC) is a type of cryptography that allows a group of parties to jointly compute results without sharing their data with each other or any trusted party [1]. By eliminating the need to trust others with sensitive data, MPC enables new collaborative applications that are currently restricted or prevented by security and privacy concerns. MPC protocols exist that can in theory be used to securely perform any computation. In recent years, these protocols have been implemented, improved, and shown to be practical for many applications [2].

MIT Lincoln Laboratory has developed a programming framework called Rapid Assembly of MPC Protocols (RAMP) that enables researchers to quickly prototype complex MPC applications. The RAMP framework has a layered architecture with a focus on modularity and generality. The lowest layer consists of MPC primitives for securely computing low-level arithmetic and Boolean operations [3], [4] on secret-shared data. On top of the primitives, we have implemented an extensive library of MPC gadgets, or data-oblivious algorithms for common computational building blocks. These gadgets abstract away complexity to allow high-level MPC functionalities to be built simply by composing a few gadgets together. Figure 1 illustrates the RAMP stack and an example of how a secure set intersection functionality decomposes into gadgets and primitives. RAMP also includes tools for evaluating the theoretical and empirical performance of MPC applications and building blocks.

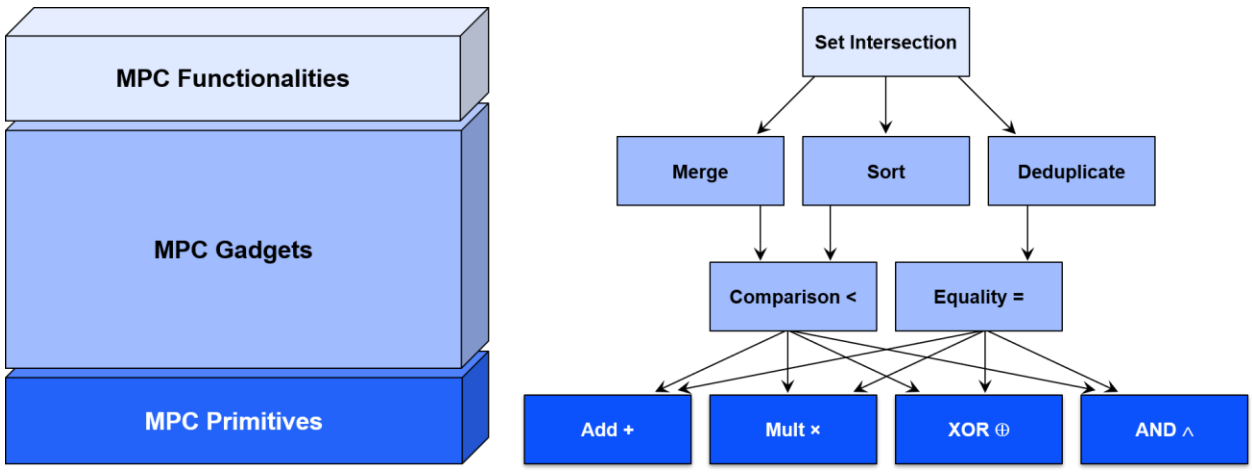


Figure 1. Rapid Assembly of MPC Protocols (RAMP) architecture and example functionality, gadgets, and primitives.

We have used the RAMP framework to prototype MPC solutions to a variety of problems. One common class of problems involves secure multi-party set operations. We have implemented secure MPC for a general version of the set intersection problem, where two or more parties with private datasets would like to learn which elements occur in at least a threshold number of the input sets, without revealing any other information. In addition, we have implemented secure computation of the similarity of two datasets. We have also combined this with secure thresholding or clustering to output information only about datasets that are highly similar. One example application area for secure set operations is collaborative cybersecurity. We have applied MPC for set intersection and similarity computations to cyber threat data such as malicious IP addresses, software vulnerabilities, and malware features.

A related problem is secure database joins. We consider the setting where multiple parties holding tables with different columns would like to merge the tables into a single table based on a sensitive unique key column, without sharing the raw data and in particular without sharing the unique keys. We have implemented MPC for secure inner and outer join operations, using algorithms similar to secure set intersection. Depending on the application, the output may be either the joined database itself (without the unique keys) or only the result of some further computation on the joined database.

Another area of growing interest is MPC for privacy-preserving machine learning. One example use case involves the application of watchlists, where one or more parties hold watchlists represented as models, and another party holds a sample and would like to learn whether that sample matches any individual in any of the watchlists, without any party learning any other information. We have prototyped MPC for probabilistic linear discriminant analysis, a machine learning approach with applications to speaker verification and face recognition. Specifically, we use MPC to securely compute the likelihood that two feature vectors are from the same speaker and determine whether the likelihood exceeds a threshold.

Beyond the examples outlined above, RAMP is a general framework that enables the development of MPC solutions for a wide variety of problems. MPC is possible for any computation and already practical for many applications. Secure multi-party computation has significant potential to enable new types of collaboration and analysis currently precluded by data sharing concerns.

References

[1] D. Evans, V. Kolesnikov and M. Rosulek, A Pragmatic Introduction to Secure Multi-Party Computation, NOW Publishers, 2018.

[2] M. Hastings, B. Hemenway, D. Noble and S. Zdancewic, "SoK: General Purpose Compilers for Secure Multi-Party Computation," in *IEEE Symposium on Security and Privacy*, 2019.

[3] M. Ben-Or, S. Goldwasser and A. Wigderson, "Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract)," in *ACM STOC*, 1988.

[4] O. Goldreich, S. Micali and A. Wigderson, "How to Play any Mental Game – A Completeness Theorem for Protocols with Honest Majority," in *ACM STOC*, 1987.

Privacy-preserving Error Resilient Record Linkage[1]

Over the past decade, private record linkage (PRL) approaches have been proposed to enable data holders to integrate records without revealing the raw information. One class is based upon secure multi-party computation (SMPC), which leverages cryptographically strong protocols. These protocols enable similarity comparisons to be performed without revealing *any* information, other than the input and output of the protocol. While secure in principle, the computationally expensive encryption techniques (e.g., homomorphic encryption, secure circuit evaluation on big data etc.) required to achieve SMPC do not scale well to very large databases with millions of records in it [1]. [2]

As an alternative, a second class of PRL, utilizes a weaker form of security based on data transformation. These transformed values, referred to as encodings, are used as inputs to PRL. It is critical to strike a balance amongst the competing priorities of accuracy, security, and efficiency when identifying an appropriate data transformation method for PRL. Our past work [1] showed that a transformation based on encoding values in a Record Level Bloom filter (RBF) is superior to other SMPC based approaches (e.g., unique hashed or obfuscated identifiers, SMPC-based equality joins, and SMPC-based similarity comparisons) in terms of efficiency and accuracy (see figure 1 for an example). Unfortunately, our past work does not provide any formal guarantees and only secure against the known attacks of the time. In our ongoing work, we are improving the privacy guarantees of the RBF schemes for PRL by leveraging trusted execution environments that are currently available on the cloud (e.g., Microsoft confidential computing cloud). This way only RBF encoded information can be sent to cloud after it is encrypted using secure public key encryption. Once the encrypted records represented as RBF is sent to the cloud, they will be only decrypted inside the trusted execution environments for record linkage purposes. Below, we provide further details on different components of the proposed tool.

Creating privacy-preserving representation of record information for linkage. In our past work [1], we created record level bloom filters (RBFs) that are tailored for private record linkage while being resistant known privacy attacks. Figure 1 illustrates the construction of RBFs for three fields for record linkage. Our evaluation based on publicly available North Carolina voter registration (NCVR) database showed that using the construction described in Figure 1, and dice coefficient based similarity metric on the RBFs, *we can achieve record linkage accuracy close to the scenarios where the entire records (even with errors) are available in plaintext. Furthermore, the known attacks where not successful against the proposed RBF generation technique.*

Leveraging cloud-based trusted execution environment tool for privacy-preserving record linkage. To securely process the RBFs that are generated for different records, we leverage trusted execution environments (TEEs) to perform computation on encrypted RBFs. TEEs allow users to execute programs securely in a manner that operating systems cannot directly observe or tamper with program execution without being detected. For example, Intel *Software Guard eX-*

[1]Contact author: **Murat Kantarcioglu, Univ. of Texas at Dallas, and DataSecTech, muratk@utdallas.edu**

[2]Obviously if there is a unique identifier such as social security number (ssn) exists, then the simple hashing approach used by some of the existing solutions could work. Unfortunately, it is known that ssn information may not be always accurate (e.g., people using their spouse ssn for insurance purposes). In this work, we target the case where there is no such uniquely identifying attribute for record linkage and/or such identifier may have errors and other information such as name and surname etc. need to be used for increased accuracy.

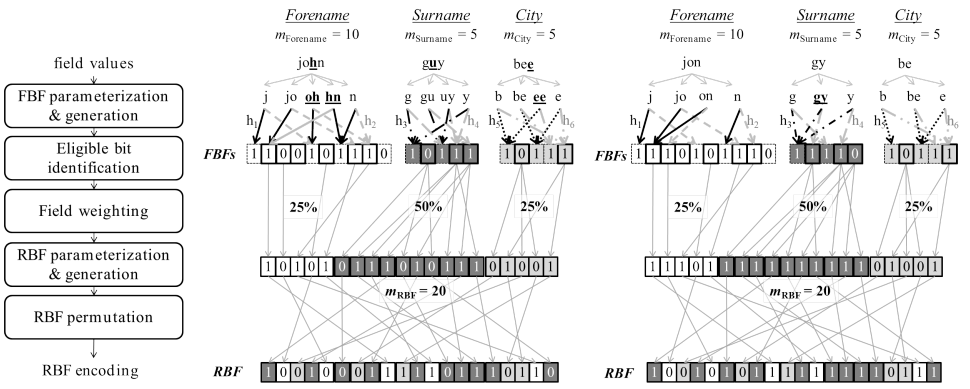


Figure 1: **RBF encoding generation.** 1) **Field Level Bloom Filter(FBF) parameterization & generation.** FBFs are sized based on the expected field length using n-grams of the string. In this example, there are two hash functions per FBF. 2) **Eligible bit identification.** More secure bits, shown with a heavy outline, are identified for inclusion in the RBF. 3) **Field weighting.** It is determined that 25% of the bits in the RBF should be drawn from *Forename*, 50% of the bits in the RBF should be drawn from *Surname*, and 25% of the bits in the RBF should be drawn from *City*. 4) **RBF parameterization & generation.** m_{RBF} is determined and eligible bits are sampled according to the field weights. 5) **RBF permutation.** Bits in the RBF are randomly permuted. The number of bits set in the RBF shown to the left and right is 11 and 15, respectively. The number of bits in the intersection is 11, yielding a Dice coefficient of 0.84. See [1] for more details.

tension (SGX) reduces the trusted code based (TCB) to a minimal set of *trusted code* (programmed by the programmer) and the *SGX processor*. Still, building a robust secure application with SGX is non-trivial due to several shortcomings of the SGX architecture. In particular, operating system can still monitor memory access patterns by the secure trusted code. Access pattern leakage can reveal a significant amount of information about encrypted data.

Currently, we are developing a privacy-preserving record linkage using TEEs on the cloud. Each site will encrypt the RBFs with the public key of the TEE that is used on the cloud, and send their encrypted RBF lists to TEE. TEE running on the cloud will decrypt the encrypted RBFs and compute dice coefficient between the RBFs for matching. As a starting point, we adopt our join algorithm [2] that is designed for TEE environments. This join algorithm can support any similarity function and is designed not leak any information to an attacker that can observe the memory access patterns of the TEE.

References

[1] E. A. Durham, M. Kantarcioglu, Y. Xue, C. Tóth, M. Kuzu, and B. A. Malin, "Composite bloom filters for secure record linkage," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2956–2968, 2014. [Online]. Available: https://doi.org/10.1109/TKDE.2013.91

[2] F. Shaon, M. Kantarcioglu, Z. Lin, and L. Khan, "Sgx-bigmatrix: A practical encrypted data analytic framework with trusted processors," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,*, 2017, pp. 1211–1228.

Multiparty Homomorphic Encryption for Privacy-Protected Linking and Querying of Watchlists

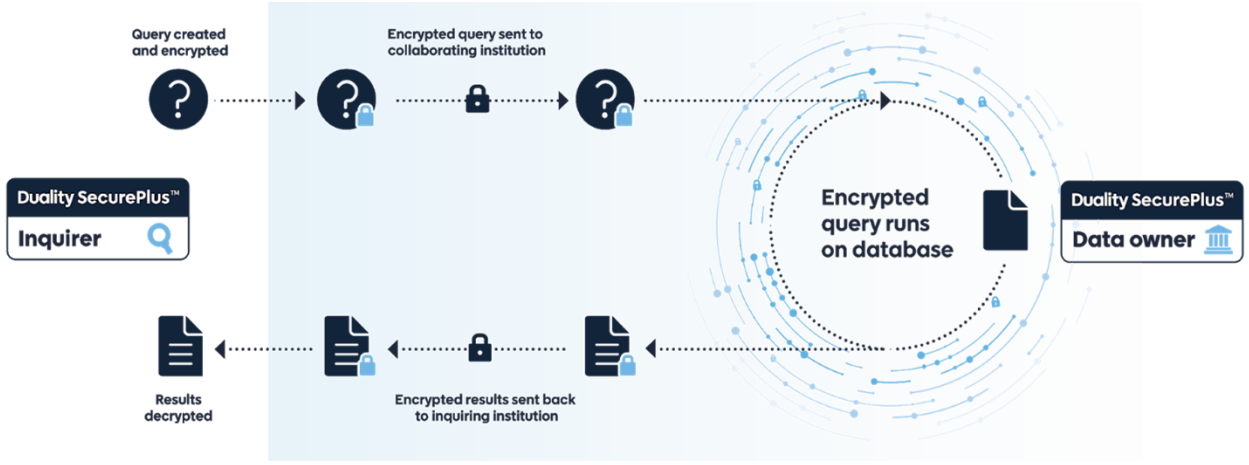
Kurt Rohloff
krohloff@dualitytech.com

Many security organizations across the United States (in federal, state, local and tribal governments and in private institutions such as airlines and other transportation agencies) are responsible for using human expertise and technologies to prevent and investigate threats. These investigations often include querying sensitive datasets such as multiple watch lists, financial transaction information, surveillance images and video and more. All of these queries and analyses to support investigations are information-based endeavors that require access to comprehensive, timely information to be effective, but are currently plagued by slow, labor intensive information gathering processes and regulations.

Streamlining the process of accessing information for inter-organizational collaboration can accelerate and improve proactive prevention and investigations and ultimately to help address national security issues such as counter-terror and counter-narcotics missions, among many others. Duality Technologies provides a Secure Querying capability to address these needs to privately query watch lists using homomorphic encryption technologies developed with support from DARPA. We discuss how these Secure Query capabilities provided by Duality Technologies can be applied and extended for quick-wins and long-term investment.

Using Fully Homomorphic Encryption for Privacy-Protected Investigations

Duality Technologies has an extensible privacy technology platform built on top of homomorphic encryption for privacy-protected information sharing to support the fight against financial crimes as outlined above. A concept of operations for this solution can be seen immediately below.



In this solution a user / investigator generates queries on sensitive data, such as a watch-list to support an investigation. To protect the privacy and anonymity of the subject of investigation, and to maintain operational security, Duality provides a capability where the investigator encrypts their query, sends the encrypted query to a query hub, potentially in a federal cloud environment, to mask the source of the query. The encrypted query is then sent on to multiple data hosts, such

as a financial institution (i.e., bank) or transportation provider (i.e., airline) to run over their data, generating an encrypted result. The data owner will never know the subject of investigation and will never know the query run on its data. The encrypted result is then sent on to the investigator who decrypts the result, thus accessing the results of their query to further their investigation. This operation could be run over multiple data sets in parallel, with aggregation and linking of the results of the queries.

Operationally, investigators can initiate privacy-protected queries which are automatically distributed to other members. Members can automatically answer these queries - without seeing any sensitive data like PII. All data and results are aggregated and anonymized, offering the utmost in privacy and security safeguards. Importantly, data never leaves a member's premises, meaning that data is decentralized and always protected.

Encryption for real-life, post-quantum applications

Duality's core technology and team grew out of a legacy of open-source homomorphic encryption technologies developed for DARPA and the DoD. All of our core encryption technologies are open-sourced, such in our PALISADE homomorphic encryption library (https://www.palisade-crypto.org) and are used in the DoD privacy technology community. We engage with multiple parts of the US government to apply this technology for national security needs, and we engage in S&T projects with the US government to generalize the capabilities, performance and scale of our offering. Homomorphic encryption schemes used by Duality Technologies are post-quantum, meaning that they are resistant to quantum computing attacks.

An S&T Vision with Quick Wins and Long-Term Value on Hard Problems

All the watch-list query capabilities are currently operational and used by Duality customers, both commercially and in government. Duality currently deploys these capabilities to support financial crime and anti-money-laundering capabilities to financial institutes, such as banks.

Additional research and development effort could easily extend Duality's capabilities as a "quick win" to provide broad generalizations of our current solutions to create, maintain, and link anonymous identities and profiles from multiple sources. Example quick-win extensions include:

- Enable investigators can homomorphically encrypt, share, and compare watchlists while preserving privacy, anonymity, and compliance
- Allow watchlist owners to query one-another based on multiple data sources with results linked dynamically on an as-needed basis.

We have early prototypes of the above functionalities we developed as a risk-reduction exercise. We also have initial capabilities to support querying on standard geospatial data frameworks.

Duality Technologies also provides broad support for longer-term high-payoff solutions to "hard problems" to integrate into our solution, inclusive of.

- Support for querying on imagery data, such as surveillance cameras or body cameras that we see as both short- and long-term value in a broader private query system.
- Support for processing of graph data, including support for social network data.

Linking Without Leaking: Private Set Intersection

Mark Blunk Paul Bunn Samuel Dittmer Steve Lu Rafail Ostrovsky

April 2022

In many research contexts where multiple parties hold data, significant insight can be gained from conducting statistical analyses across disjoint datasets; however, privacy considerations frequently render collaboration and data-sharing infeasible. Such considerations are especially prohibitive in highly regulated sectors that routinely generate and obtain data of a particularly sensitive nature, including healthcare and social services, financial services, socio-economic research, the military, and national security. At the same time, the sensitivity of these data is frequently highly correlated with the value of information that can be derived from analyzing them, and as a result of this tension, a significant amount of social benefit is left unrealized.

One significant area of application is in the construction and application of watch lists. Suppose a transportation company (say, an airline or a train company) holds a dataset containing a passenger manifest, while a government agency (say, the TSA) holds a dataset containing a list of people on a watch list. In this setting, the transportation service party wishes to verify whether any passengers are on the government's watch list.

In the specific case of the DHS No Fly List, this problem is currently solved by transferring all of the passenger lists to the government agency, which then compares this list to their internal database. However, as privacy concerns grow, this arrangement may not always be tenable. The solution to this watch list problem with the broadest privacy protections is a powerful technology called Private Set Intersection (or PSI). This allows the transportation company and the government agency to learn which passengers, if any, are on the government agency's watch list, while ensuring that the government agency learns nothing about the other passengers, and the transportation company learns nothing about who is on the list (besides the names of the specific passengers).

The key feature of this situation that makes PSI applicable is that, once a passenger on both the manifest and the watch list is identified, the privacy concerns are overruled by more pressing considerations. In the case of the No Fly List, the more pressing consideration is a terrorist threat, but this same pattern appears in many other contexts, making PSI a valuable tool.

For example, if an internet user wishes to learn if any of their passwords have been hacked, their list of passwords becomes the "flight manifest", and the list of known hacked passwords becomes the "government watch list". Here, the user does not wish anyone to have access to their passwords, and the company maintaining the list of known hacked passwords does not want to spread that list around to any malicious actors. But, for the particular case where the user's password is in the list of known hacked passwords, these privacy concerns are overruled by more pressing considerations - namely, the user wants to change their password immediately, and so must know which passwords to change.

As another example, the "flight manifest" could represent some list of persons observed near the scene of a crime during a local police investigation, while the "watch list" could represent another state or federal database of persons of interest in previous cases. The privacy concerns here might be regulatory in nature, for example, there could be rules against sharing certain federal databases too widely, and/or rules against sharing data from local police departments without probable cause. Both of these concerns would be overruled precisely when the combined evidence from both lists

showed the person warranted further investigation. More broadly, many problems of information sharing between government agencies could be addressed in this way - whenever there are a series of indicators that are insignificant individually, but deserve attention if all indicators are present simultaneously.

At Stealth Software Technologies, Inc., as part of a recently completed Phase I Small Business grant funded by NIST, we have developed a prototype tool for PSI that operates in the specific setting of many parties interacting with a semi-trusted central coordinator computing their intersections pairwise, that is, each party learns their intersection with each other party. Modifying the example of a local law enforcement department above, this tool would be helpful when the departments of several nearby municipalities wished to share evidence with each other in the investigation of (for example) a string of possibly related murders. We have tested our implemented prototype extensively in local settings and in multi-server deployments on Amazon Web Services, and plan on further developing and optimizing this work once we are approved for the next phase in this grant.

There is a variant of PSI, *PSI with payloads* that is also worth highlighting. In this scenario the identifiers each have some associated data, and the protocol performs a computation on the data of the matched identifiers. This tool is simultaneously more powerful, since additional data statistics can be extracted, and more privacy-preserving, since now only the final statistic, and not the intermediate records, are revealed.

For example, street cameras on either side of a busy city bridge could be used to compute the average time required to cross the bridge by matching license plates of people driving onto and off of the bridge. The act of driving onto and off of a bridge is not, of course, a pressing consideration that warrants disclosing an individual's identity. However, when instead the protocol computes only the average time difference across all people who drove on the bridge, useful traffic information can still be extracted without disclosing anyone's identity. Such computations can be thought of as a Private/Secure analogue of a SQL join, followed by an aggregate computation on the merged dataset.

In projects funded by DARPA and Arnold Ventures, we have designed systems that allow multiple data owners to participate in such computations. In particular, they can compute cross-tabulations of counts and averages. Our implementations have several components to enhance their usability by non-experts in Cryptography. They are designed to integrate with modern RDBMS systems, to operate on data in a streaming fashion, enabling the protocol to easily scale to large input datasets, and make use of recent development in secure merge to optimize the overall algorithmic complexity of the protocol. Both implementations have been tested on production servers maintained by our respective partners (educational and state government entities) on these projects.

In summary, we feel that the field of Private Set Intersection presents a significant opportunity to enable collaborative analyses among organizations that are currently unable to due to privacy and security considerations. Our company and its consultants have experience in both research and implementation development in this field, are excited to discuss any opportunities to use this technology to solve real-world problems.