

Hawk: Accurate and Fast Privacy-Preserving Machine Learning Using Secure Lookup Table Computation

Hamza Saleem

University of Southern California
hsaleem@usc.edu

Muhammad Naveed

University of Southern California
mnaveed@usc.edu

Amir Ziashahabi

University of Southern California
ziashaha@usc.edu

Salman Avestimehr

University of Southern California
avestime@usc.edu

ABSTRACT

Training machine learning models on data from multiple entities without direct data sharing can unlock applications otherwise hindered by business, legal, or ethical constraints. In this work, we design and implement new privacy-preserving machine learning protocols for logistic regression and neural network models. We adopt a two-server model where data owners secret-share their data between two servers that train and evaluate the model on the joint data. A significant source of inefficiency and inaccuracy in existing methods arises from using Yao’s garbled circuits to compute non-linear activation functions. We propose new methods for computing non-linear functions based on secret-shared lookup tables, offering both computational efficiency and improved accuracy.

Beyond introducing leakage-free techniques, we initiate the exploration of relaxed security measures for privacy-preserving machine learning. Instead of claiming that the servers gain no knowledge during the computation, we contend that while some information is revealed about access patterns to lookup tables, it maintains ϵ - $d\chi$ -privacy. Leveraging this relaxation significantly reduces the computational resources needed for training. We present new cryptographic protocols tailored to this relaxed security paradigm and define and analyze the leakage. Our evaluations show that our logistic regression protocol is up to $9\times$ faster, and the neural network training is up to $688\times$ faster than SecureML [58]. Notably, our neural network achieves an accuracy of 96.6% on MNIST in 15 epochs, outperforming prior benchmarks [58, 76] that capped at 93.4% using the same architecture.

KEYWORDS

secure multi-party computation, privacy-preserving ML

1 INTRODUCTION

Machine learning (ML) has become an indispensable tool in various domains, from advertising and healthcare to finance and retail, enabling the training of predictive models. The accuracy of these models is significantly enhanced when trained on vast datasets aggregated from diverse sources. However, numerous challenges, such as privacy concerns, stringent regulations, and competitive

business landscapes, often hinder collaborative data-sharing endeavors.

Privacy-Preserving Machine Learning (PPML) using secure computation allows multiple entities to train models on their aggregated data without revealing their individual datasets. In this paradigm, data owners distribute their private data among several non-colluding servers using secret-sharing techniques. These servers then perform training on the combined data without gaining insights into the individual datasets.

Past works on PPML have focused on training logistic regression models [58, 81] and neural networks [26, 57, 58, 76]. A significant bottleneck in existing work is using expensive protocols, such as Yao’s garbled circuits, for computing non-linear activation functions. Furthermore, prior works [58, 76] resort to non-standard or approximate activation functions, resulting in accuracy loss.

In this work, we present two novel protocols $\text{HAWK}_{\text{Single}}$ and $\text{HAWK}_{\text{Multi}}$, designed for accurate and efficient computation of activation functions and their derivatives. These protocols leverage secret shared precomputed lookup tables. We begin by introducing the $\text{HAWK}_{\text{Single}}$ protocol, which offers plaintext accuracy and robust security in a semi-honest adversary model. Its primary limitation, however, is its reliance on consuming one lookup table for each function computation, necessitating a vast number of such tables. To mitigate this limitation, we subsequently introduce $\text{HAWK}_{\text{Multi}}$, enabling lookup table reusability. The $\text{HAWK}_{\text{Multi}}$ protocol poses several security and efficiency challenges, which we further address to achieve a fast and secure protocol. Although methods such as Oblivious RAM (ORAM) [36] can be employed to eliminate information leakage in our context, we opt for a slightly relaxed security model to prioritize performance improvements. Our $\text{HAWK}_{\text{Multi}}$ protocol permits table reuse but at the cost of revealing limited information about the access patterns. We rigorously analyze the leakage of the $\text{HAWK}_{\text{Multi}}$ protocol and prove that it preserves ϵ - $d\chi$ -privacy for the leaked access patterns.

Building on $\text{HAWK}_{\text{Single}}$ and $\text{HAWK}_{\text{Multi}}$ we develop efficient PPML protocols tailored for logistic regression and neural network training in a two-server secure computation environment. Our evaluations, conducted on Amazon EC2, showcase the performance gains of our system.

1.1 Our Contributions

Computing Standard Activation Functions. We propose a novel method to accurately compute activation functions in a secure computation setting and achieve accuracy similar to plaintext training.

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license visit <https://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Proceedings on Privacy Enhancing Technologies 2024(3), 42–58
© 2024 Copyright held by the owner/author(s).
<https://doi.org/10.56553/popets-2024-0066>

Past methods either use non-standard activations or approximate them. As shown in Figure 9, using non-standard activation functions may cause accuracy to plummet to random guess levels, potentially affecting convergence. Hence, using standard activation functions is crucial to harness machine learning advancements rather than relying on makeshift functions for secure computation.

Generic Method for Precise Univariate Function Computation. We introduce a universal method for securely computing any univariate function. It can be directly used to compute any univariate activation function, which constitutes most of the activation functions used in all but the last layer of a neural network. Our method computes activations much faster than the state-of-the-art. This method creates secret shared lookup tables for all possible inputs; we use recent advances in low-precision machine learning to confine the size of the lookup tables.

Precise Computation of Multivariate Functions. We demonstrate the application of our method to compute multivariate activations, notably Softmax. We highlight that this is the first method to compute Softmax in a secure computation setting accurately.

Beyond Activation Functions. While we demonstrate the applicability of our protocols for activation function computation, their potential extends far beyond these specific functions. The underlying principle of precomputed lookup tables can be applied to a wide range of MPC applications where the input domain can be bounded. For example, in private set intersection [44], our protocol can enable efficient comparison of elements in private sets. Similarly, secure auctions [12] can benefit from lookup tables to ensure privacy while evaluating bids. Furthermore, our protocols can be used to securely compute histograms [55], enabling private data analysis. These applications highlight the versatility of our protocols for MPC across various domains.

An Alternate Representation for DReLU Activation. We present an alternate representation of the DReLU activation function for fixed-point numbers, which allows us to reduce the entries in the lookup table for DReLU by a factor of 2^{11} .

Relaxed Security Setting. In addition to introducing leakage-free methods using Hawk_{Single}, we initiate the study of a new security model for PPML. This relaxed model allows some information leakage about access patterns, but the leakage is proven to preserve $d_{\mathcal{X}}$ -privacy. We leverage this model to develop Hawk_{Multi} protocol to significantly reduce computational resources while training.

Experimental results. We implement our framework and perform various experiments for training and inference of logistic regression and neural networks using six different datasets in both LAN and WAN settings. Section 8 shows that our protocols are significantly faster than state-of-the-art in a two-party setting and achieve accuracy similar to plaintext training.

2 RELATED WORK

This section overviews prior studies across fields such as access pattern privacy, PPML, and secure MPC pertinent to our work.

Access Pattern Privacy literature focuses on achieving full or partial obliviousness during data retrieval. ORAM [36] algorithms provide complete obliviousness by ensuring that the physical access patterns for any two equal-length logical patterns are indistinguishable. While ORAM addresses access pattern leakage and has been

extensively studied in both client-server [5, 36, 62, 67, 70] and multi-party [28, 37, 62, 79, 85] settings, it carries an inevitable overhead of $\Omega(\log(n))$ [15, 36, 50].

A natural approach for circumventing this overhead is relaxing full obliviousness for a weaker privacy notion, such as differential privacy. Differential privacy introduces a neighborhood concept and ensures closeness between neighboring access sequences rather than all sequences. Despite the relaxation, [61] show that the $\Omega(\log(n))$ lower bound also applies to DP-ORAM [75] structures. Nonetheless, various studies have employed differentially private access patterns to offer performance enhancements for specific algorithms, such as searchable symmetric encryption (SSE) [66], sorting and merging of lists [20], histograms [2], and graph-parallel computation [55]. The performance enhancement for these algorithms stems from the fact that they generally rely on a full memory scan. Therefore, the problem of access pattern privacy is reduced to hiding the number of accesses for each element. For lookup tables, where only a limited number of elements are accessed, these algorithms fail to deliver an efficient solution.

PPML Using Secure MPC. Previous MPC studies have explored decision trees [52], K-means clustering [16, 45], SVM [72, 84], and regression techniques [30, 65, 69, 80]. However, these works incur high efficiency overheads and lack implementation. SecureML [58] provides protocols for secure training and inference of linear regression, logistic regression, and neural networks in a two-server setting. They present alternate functions to approximate the non-linear activations (Sigmoid, ReLU, and Softmax), which are evaluated using garbled circuits. However, garbled circuits have a high communication overhead, and approximating activation functions reduces the model’s accuracy. QUOTIENT [1] presents protocols for neural network training in a two-server setting, utilizing quantization for various network components. However, using garbled circuits and oblivious transfer produces high overheads.

SecureNN [76] and ABY³ [57] present privacy-preserving neural network training protocols in a three-server honest majority setting. Like SecureML, they approximate Softmax, which reduces the model accuracy, as discussed in Section 8.4. Falcon [77] combines techniques from SecureNN [76] and ABY³ [57] to improve the training performance. Our two-server protocol, as shown in Section 8, outperforms these in efficiency. Another pertinent work in the three-server setting is Pika [73], which adopts a lookup table-centric approach to computing non-linear functions. Pika relies on Private Information Retrieval (PIR) and Function Secret Sharing (FSS) [14] to hide access patterns. Despite providing full obliviousness, their protocol has an overhead of $O(n)$ stemming from FFS key generation, FFS evaluation, and dot product operation with all lookup table elements. In fact, $\Omega(n)$ overhead is intrinsic to PIR [10]. Pika only provides performance evaluations for generic functions and inference tasks. We compare our protocols with Pika for sigmoid function computation in Appendix F.

A parallel line of work focuses on secure neural network inference. A party or a group of parties evaluates the model on a client’s private data, keeping the model private. Gilad-Barach et al. [29] employ HE for secure inference, albeit with accuracy compromises. MiniONN [53] optimizes the secure inference protocols of SecureML [58] by reducing the offline cost of matrix multiplication. Using a third-party dealer, Chameleon [64] removes the oblivious

transfer protocols required for multiplication. Gazelle [46] optimizes linear operations using specialized packing schemes for HE. Baccarini et al. [7] present ring-based protocols for MPC operations applied to NN inference. SIRNN [63] presents a lookup table based protocol that uses digit decomposition with oblivious transfers to compute non-linear functions efficiently. We compare our protocols with SIRNN for sigmoid function computation in Appendix F.

Other PPML Approaches. Shokri and Shmatikov [68] explore privacy-preserving neural network training with horizontally partitioned data. The servers train separately and share only parameter updates, using differential privacy [31] to minimize leakage. Wu et al.[81] and Kim et al.[49] employ homomorphic encryption for logistic regression, approximating the logistic function with polynomials. The complexity is exponential in the polynomial degree, and the approximation reduces the model accuracy. Crawford et al. [22] use fully homomorphic encryption to train logistic regression models using encrypted lookup tables. However, the approach incurs high overhead, and using low-precision numbers reduces the model’s accuracy. In contrast, our protocols are orders of magnitude more efficient and achieve accuracy similar to plaintext training.

We primarily compare our techniques with SecureML [58] as this is the only work we found for logistic regression and neural network training in a two-server client-aided setting. While recent works exist [1, 57, 76, 77], their settings differ, making direct comparisons challenging. We compare our online cost with theirs as they lack an offline phase. We further remark that the idea of employing lookup tables for function evaluation is not novel with past works [14, 24, 27, 48, 63] suggesting circuit and FSS-based protocols. However, these works necessitate representing functions as either boolean or arithmetic circuits, with each gate evaluated using a lookup table. It is non-trivial to directly represent functions like exponential, logarithm, etc, used in ML activation, as circuits. This complexity typically leads to reliance on circuit-friendly approximations, such as the Taylor series, which compromises accuracy and introduces inefficiencies (see Figure 9). Our lookup table based protocols are fundamentally different as they enable direct computation of ML activations without resorting to circuit approximations. Furthermore, our main contribution is the HAWK_{Multi} protocol that allows for the reuse of a lookup table by leveraging metric differentially private access patterns, missing in prior research.

3 PRELIMINARIES

3.1 Machine Learning

Logistic Regression is a binary classification algorithm commonly, often used for tasks like medical diagnosis. Given n training examples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^D$ with binary labels y_1, y_2, \dots, y_n , the goal is to learn a coefficient vector $\mathbf{w} \in \mathbb{R}^D$ which minimizes the distance between $g(\mathbf{x}_i) = f(\mathbf{x}_i \cdot \mathbf{w})$ and the true label y_i . The *activation function* f is used to bound the dot product of \mathbf{x}_i and \mathbf{w} between 0 and 1 for binary classification. The Sigmoid function is typically chosen for f , defined as: $f(z) = \frac{1}{1+e^{-z}}$. To learn \mathbf{w} , we define a cost function $C(\mathbf{w})$ and find \mathbf{w} using the optimization $\arg \min_{\mathbf{w}} C(\mathbf{w})$. We use the cross-entropy cost function defined as: $C_i(\mathbf{w}) = -y \log f(\mathbf{x}_i \cdot \mathbf{w}) - (1 - y) \log(1 - f(\mathbf{x}_i \cdot \mathbf{w}))$, with the overall cost being the average across all examples: $C(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n C_i(\mathbf{w})$.

Neural Networks extend logistic regression to handle more complex tasks. A neural network consists of l layers where each layer i contains d_i neurons. The output of each neuron is computed by applying a non-linear activation to the dot product of its coefficient vector \mathbf{w} and the input \mathbf{x} . Activation functions, like the ReLU: $f(z) = \max(0, z)$, are commonly used. The first layer receives inputs from the dataset, with each subsequent layer performing calculations and forwarding the output to the next layer using ‘forward propagation’. For classification, the probability of the input example belonging to each class is computed by applying the Softmax function: $f(z_i) = \frac{e^{z_i}}{\sum_{i=1}^{d_i} e^{z_i}}$ to each neuron in the output layer, where d_i denotes the total neurons in the output layer. The Softmax output is a probability value between 0 and 1 for each class, summing up to 1. To train the network, we apply the Stochastic Gradient Descent (SGD) algorithm and update the coefficients of all neurons recursively, starting from the last layer and moving towards the first layer in a ‘backward propagation’ step. For further details on related ML concepts like SGD and batching, refer to Appendix A.

3.2 Secure Computation

Secure computation allows parties to compute a joint function of their data while keeping their data private.

Preprocessing Model of Secure Computation divides a protocol into two phases: a data-independent offline phase used to create correlated randomness and an online phase. This model allows most of the heavy computation to be delegated to the offline preprocessing phase, rendering the online phase more computationally efficient.

Additive Secret-Sharing. In a secure two-party computation (2PC) setting, secret values are additively secret-shared between the parties P_0 and P_1 . To share a secret $x \in \mathbb{Z}_n$, a random $r \in \mathbb{Z}_n$ is generated. The share for the first party is computed as $\langle x \rangle_0 = r$, while the second party’s share is $\langle x \rangle_1 = x - r \pmod{n}$. To reconstruct x , one party sends its share to the other party, which then computes $x = \langle x \rangle_0 + \langle x \rangle_1 \pmod{n}$, denoted as $Rec(x_0, x_1)$. We use $\langle x \rangle_i$ to denote P_i ’s share of x .

Secure Addition. The parties can non-interactively add two secret-shared values x and y where each P_i computes $\langle z \rangle_i = \langle x \rangle_i + \langle y \rangle_i \pmod{n}$ as their share of the sum.

Secure Multiplication. To multiply two secret-shared values x and y , the parties use Beaver’s multiplication triples [8]. Initially, the parties possess shares of a, b , and c where a and b are random values in \mathbb{Z}_n and $c = ab \pmod{n}$. Each party P_i computes $\langle d \rangle_i = \langle x \rangle_i - \langle a \rangle_i \pmod{n}$ and $\langle e \rangle_i = \langle y \rangle_i - \langle b \rangle_i \pmod{n}$ and compute $d = Rec(d_0, d_1)$ and $e = Rec(e_0, e_1)$ to reconstruct d and e . Each party P_i then computes $\langle z \rangle_i = \langle x \rangle_i e + \langle y \rangle_i d + \langle c \rangle_i - ide$ as their share of the product.

3.3 Metric Differential Privacy (d_{χ} -privacy)

DEFINITION 1. A randomized algorithm \mathcal{L} is ϵ -differentially private if for all datasets D_1 and D_2 differing in a single value and for all $S \subset Range(\mathcal{L})$: $Pr[\mathcal{L}(D_1) \in S] \leq e^\epsilon Pr[\mathcal{L}(D_2) \in S]$, where the probability space is over the coin flips of the mechanism \mathcal{L} .

While the above definition captures differential privacy in a centralized setting, there are scenarios where data privacy at the

individual level is paramount. Local Differential Privacy (LDP) obfuscates individual data records when clients distrust the central curator, altering data before statistical analysis.

DEFINITION 2. A randomized algorithm \mathcal{L} provides ϵ -LDP if for each pair of inputs x and x' and for all $S \subset \text{Range}(\mathcal{L})$: $\Pr[\mathcal{L}(x) \in S] \leq e^\epsilon \Pr[\mathcal{L}(x') \in S]$.

In this work, we adopt a natural generalization of the conventional differential privacy [31] presented by Chatzikokolakis et al. [21]. This extension is particularly relevant when the domain \mathcal{X} is viewed as a metric space. Termed as metric differential privacy or $d_{\mathcal{X}}$ -privacy, this approach is adept at modeling situations where indistinguishability is best expressed through a metric between secret inputs. Geo-Indistinguishability is presented as a canonical example for $d_{\mathcal{X}}$ -privacy [3, 21], where a user wishes to share their location with a service to receive nearby restaurant recommendations. Instead of disclosing the precise coordinates, the user can employ $d_{\mathcal{X}}$ -privacy to share an *approximated location* to receive relevant suggestions without compromising their exact whereabouts.

The input domain \mathcal{X} for an activation function can be viewed as a metric space. For instance, if for some $x, x_1, x_2 \in \mathcal{X}$, $|x - x_1| < |x - x_2|$, then usually, $f(x_1)$ provides a better approximation of $f(x)$ than $f(x_2)$. Mechanisms from standard or local DP can be adapted to suit $d_{\mathcal{X}}$ -privacy by choosing a fitting metric for the confidential data. We leverage local $d_{\mathcal{X}}$ -privacy to obfuscate each individual table lookup in our Hawk_{Multi} protocol, ensuring enhanced privacy from computing servers.

DEFINITION 3. A randomized algorithm \mathcal{L} provides local ϵ - $d_{\mathcal{X}}$ -privacy if for each pair of inputs x and x' and for all $S \subset \text{Range}(\mathcal{L})$: $\Pr[\mathcal{L}(x) \in S] \leq e^{\epsilon \cdot d_{\mathcal{X}}(x, x')} \Pr[\mathcal{L}(x') \in S]$, where $d_{\mathcal{X}}(\cdot, \cdot)$ is a distance metric.

We define the distance metric as the absolute difference between two points, i.e., $d_{\mathcal{X}}(x, x') = |x - x'|$. We choose $d_{\mathcal{X}}$ -privacy as the noising mechanism to obfuscate table lookups while ensuring the noised value remains close to the original. This proximity is crucial for maintaining model accuracy. By definition, $d_{\mathcal{X}}$ -privacy offers a stronger privacy guarantee for inputs that are close to each other due to similar output distributions after adding noise. We can achieve a stronger guarantee for distant input values by adding large noise but at the cost of utility. Our protocol aims to balance privacy, accuracy, and efficiency. Opting for stronger privacy guarantees and higher accuracy would require more lookup tables, each with a smaller privacy budget but at a higher offline phase and storage costs. Conversely, optimizing for privacy and efficiency would necessitate adding more noise to each lookup, potentially compromising accuracy. Lastly, for leakage analysis, it is common in related studies [55] to focus on access pattern leakage. Analyzing potential leaks for the underlying input data tends to be quite complex.

Differential privacy employs various mechanisms to introduce controlled randomness to query results. In general, mechanisms designed for differential privacy can be extended to $d_{\mathcal{X}}$ -privacy. The geometric mechanism adds integer-valued noise from a symmetric geometric distribution to the output of f , ensuring ϵ - $d_{\mathcal{X}}$ -privacy.

DEFINITION 4. Let $f : \mathcal{X} \rightarrow \mathbb{Z}$ and let γ be drawn from a symmetric geometric distribution with the probability mass function: $G_x = \frac{1-e^{-\epsilon}}{1+e^{-\epsilon}} \cdot e^{-\epsilon \cdot d_{\mathcal{X}}(x, x')}$. Then, $\mathcal{A}(\cdot) = f(\cdot) + \gamma$ is ϵ - $d_{\mathcal{X}}$ -private.

We provide the proof in Appendix C.

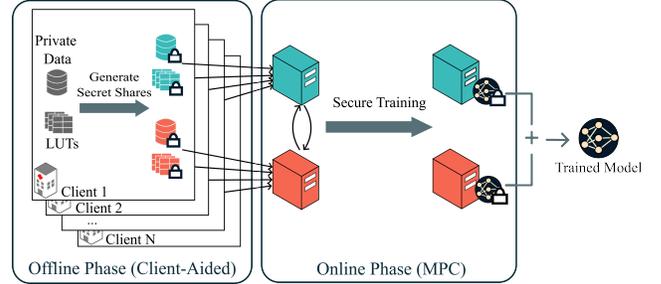


Figure 1: Overview of our PPML setup.

4 SECURITY MODEL

We consider a setting where clients C_0, \dots, C_{n-1} want to train ML models on their joint data. The data can be horizontally or vertically partitioned among the clients. The clients outsource the computation to two untrusted but non-colluding servers, P_0 and P_1 . The clients secret share the input data between P_0 and P_1 that employ secure 2PC techniques to train ML models on this data. Figure 1 depicts an overview of the setup.

Client-Aided Offline Phase. We use a client-aided offline phase. Since clients must initially upload their data to the servers, we further mandate them to generate and transmit random numbers and lookup tables, which are later utilized in the online phase. This strategy considerably enhances the efficiency of the offline phase by imposing only a marginal additional load on the clients, who are uninvolved post this phase. An alternative is the server-aided offline phase, where the computing servers use homomorphic encryption and oblivious transfer based techniques to generate required randomness. However, as noted by past work [58], this method is a significant bottleneck for PPML due to extensive cryptographic tasks. Aiming for practical PPML protocols, we adopt a client-aided offline phase, which provides a significant speed up as seen in Table 5. The only limitation of a client-aided offline phase is that the participating clients can not collude with the computing servers. This ‘‘Client-Aided Setting’’ has been formalized and adopted in several past works on PPML [34, 47, 58]. It is worth noting that each client supplies randomness and lookup tables exclusively for their specific batch of training data. We refer to the client executing the offline phase as a Client – Aide.

4.1 Security Definition

We consider a semi-honest adversary \mathcal{A} , which adheres to protocol specifications but tries to obtain information about the clients’ data. We assume \mathcal{A} can corrupt at most one of the two servers and any client subset, provided corrupt clients don’t collude with servers. The security definition requires that adversary \mathcal{A} only learns the data of corrupted clients and the final trained model but does not learn the data of any honest client.

We prove the security of our protocols using the simulation paradigm [17, 19, 35]. The universal composition framework [19] facilitates security arguments for diverse protocol compositions. End-to-end training security follows from the composition of underlying protocols. Security is modeled using two interactions: a real interaction where the servers run a two-party protocol Π in the presence of an adversary \mathcal{A} as defined above and an *environment* \mathcal{Z} , and an ideal interaction where the servers send the inputs to a functionality \mathcal{F} that performs the desired computation truthfully.

DEFINITION 5. *A protocol Π securely realizes \mathcal{F} if, for every adversary \mathcal{A} acting in the real world $\text{Real}[\mathcal{Z}, \mathcal{A}, \Pi, \lambda]$, there is an adversary \mathcal{S} in the ideal world $\text{Ideal}[\mathcal{Z}, \mathcal{S}, \mathcal{F}, \lambda]$, such that the probability of \mathcal{A} succeeding in the real world interaction and the probability of \mathcal{S} succeeding in the ideal world interaction, given the security parameter λ , are related as:*

$$|Pr[\text{Real}[\mathcal{Z}, \mathcal{A}, \Pi, \lambda] = 1] - Pr[\text{Ideal}[\mathcal{Z}, \mathcal{S}, \mathcal{F}, \lambda] = 1]| = \text{negl}(\lambda)$$

We introduce two important modifications to the security definition to establish security for $\text{Hawk}_{\text{Multi}}$ protocol, which permits access pattern leakages. Firstly, we allow some leakage in the ideal world to mirror the information learned by the adversary in the real world by observing the lookup table accesses. The leakage function is a randomized function of the access patterns. Secondly, we impose an extra condition requiring this leakage function to maintain $d_{\mathcal{X}}$ -privacy. In the ideal interaction, each party P_i forwards its input to the ideal functionality. The ideal functionality reconstructs the input received x_1, x_2, \dots, x_n , computes the output $f(x_1, x_2, \dots, x_n)$, and sends $(f(x_1, x_2, \dots, x_n))_0$ to P_0 and $(f(x_1, x_2, \dots, x_n))_1$ to P_1 . Concurrently, the ideal functionality computes a leakage function \mathcal{L} of the input and sends \mathcal{L} to the simulator \mathcal{S} . Our protocols operate in a hybrid world with access to secure, ideal functionalities, which we substitute with secure computation techniques in our application. Per Canetti's classic result [18], proving security involves treating these as trusted functionality calls.

DEFINITION 6. *A protocol Π securely realizes \mathcal{F} with leakage \mathcal{L} and (λ, ϵ) -security, while making calls to an ideal functionality \mathcal{G} , if, \mathcal{L} is ϵ - $d_{\mathcal{X}}$ -private, and for every adversary \mathcal{A} acting in the \mathcal{G} -hybrid world $\text{Hybrid}[\mathcal{G}, \mathcal{Z}, \mathcal{A}, \Pi, (X_0, X_1), \lambda]$, there is an adversary \mathcal{S} in the ideal world $\text{Ideal}[\mathcal{Z}, \mathcal{S}(\mathcal{L}(X)), \mathcal{F}, (X_0, X_1), \lambda]$, such that the probability of \mathcal{A} succeeding in the hybrid world interaction and the probability of \mathcal{S} succeeding in the ideal world interaction, on valid inputs X_0 and X_1 , are related as:*

$$|Pr[\text{Hybrid}[\mathcal{G}, \mathcal{Z}, \mathcal{A}, \Pi, (X_0, X_1), \lambda] = 1] -$$

$$Pr[\text{Ideal}[\mathcal{Z}, \mathcal{S}(\mathcal{L}(X)), \mathcal{F}, (X_0, X_1), \lambda] = 1]| = \text{negl}(\lambda)$$

5 ACCELERATING SECURE COMPUTATION WITH LOOKUP TABLES

We propose two novel secure computation protocols that employ lookup tables to evaluate a specific univariate function, $y = f(x)$. These protocols enable the secure computation of non-linear activations. Furthermore, we detail how to extend our method to devise specialized protocols for certain multivariate functions. Specifically, we present a protocol for computing the Softmax function using univariate functions. We use the preprocessing model of secure computation discussed in Section 3.2.

Algorithm 1 Table Generation: $\Pi_{\text{HawkSingle}}^{\text{offline}}(\text{Client-Aide})$

Input: Total number of lookup tables m , function $f : X \rightarrow Y$, and a cryptographic hash function H .

Output: P_i gets lookup tables $L_i^{f,c}$ for $0 \leq c < m$.

1 Client – Aide sends a random key k_i to P_i .

2 For $0 \leq c \leq m - 1$:

2.1 For all $x \in X$:

a Client – Aide computes additive secret shares $\langle y \rangle_0$ and $\langle y \rangle_1$ of $y = f(x)$.

b Client – Aide adds the *key* : $H(x + H(k_1|c))$, *value* : $\langle y \rangle_0$ pair in $L_0^{f,c}$.

c Client – Aide adds the *key* : $H(x + H(k_0|c))$, *value* : $\langle y \rangle_1$ pair in $L_1^{f,c}$.

2.2 Client – Aide randomly permutes all key value pairs in $L_0^{f,c}$ and $L_1^{f,c}$.

3 Client – Aide sends all $L_0^{f,c}$ for $0 \leq c < m$ to P_0 .

4 Client – Aide sends all $L_1^{f,c}$ for $0 \leq c < m$ to P_1 .

Algorithm 2 Query Table: $\Pi_{\text{HawkSingle}}^{\text{online}}(P_0, P_1)$

Input: For $i \in \{0, 1\}$, P_i holds key k_i , $\langle x \rangle_i$, and a cryptographic hash function H .

Output: P_i gets secret share $\langle y \rangle_i$ of $y = f(x)$.

Initialization: P_i receives lookup tables $L_i^{f,c}$ for all $0 \leq c < m$ from Client – Aide. P_0 and P_1 initialize $c = 0$.

1: P_i computes $\langle x \rangle_i + H(k_i|c)$ and sends it to P_{1-i} .

2: P_i computes $H(x + H(k_{1-i}|c)) = H(\langle x \rangle_i + \langle x \rangle_{1-i} + H(k_{1-i}|c))$.

3: P_i retrieves the value: $\langle y \rangle_i$ corresponding to the key: $H(x + H(k_{1-i}|c))$ from $L_i^{f,c}$.

4: P_0 and P_1 increment $c = c + 1$.

5: P_i outputs $\langle y \rangle_i$.

During the offline phase, for a party P_i , a lookup table L_i^f can be constructed, where $f : X \rightarrow Y$ is a univariate function and X and Y are the domain and range of f , respectively. Each entry in L_i^f is of the form *key* : $\text{PRF}(K, x)$, *value* : $\langle f(x) \rangle_i$ for every $x \in X$. Here $\text{PRF}(K, \cdot)$ denotes a pseudo-random function with key K . In the online phase, P_i holds a secret share $\langle x \rangle_i$ and seeks to determine $\langle f(x) \rangle_i$. As $\langle x \rangle_i$ is a secret share generated in the online phase, it cannot directly serve as a key for the value $\langle f(x) \rangle_i$. The key for this value must be deterministically derived from x . During the online phase, computation parties collaborate to securely compute $\text{PRF}(K, x)$ from the shares $\langle x \rangle_i$ and use it to retrieve $\langle f(x) \rangle_i$ from the lookup table L_i^f . To this end, we propose two protocols $\text{Hawk}_{\text{Single}}$ and $\text{Hawk}_{\text{Multi}}$. While $\text{Hawk}_{\text{Single}}$ provides perfect security, it comes with a higher offline cost and increased storage overhead. In contrast, $\text{Hawk}_{\text{Multi}}$ permits table reuse, substantially reducing these overheads, but it does introduce some access pattern leakage, which is proved to be $d_{\mathcal{X}}$ -private.

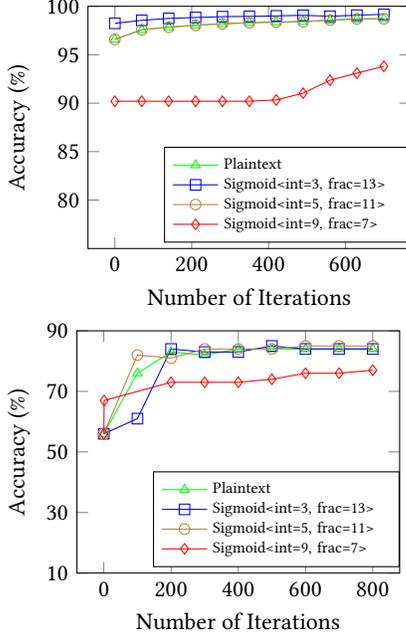


Figure 2: a) MNIST ($|B| = 128$) b) Arcene ($|B| = 100$)
Accuracy comparison of plaintext training with our PPLR protocol considering different bit representations for the Sigmoid function.

5.1 $\text{HAWK}_{\text{Single}}$: Single Use Lookup Table Protocol

In the online phase of $\text{HAWK}_{\text{Single}}$, each computation of f consumes an entire table. We generate a distinct table for every lookup during the offline phase. The offline and online phases are detailed in Algorithms 1 and 2, respectively. Here, c ensures each table has distinct keys. For practical utility, the size of the table for $\text{HAWK}_{\text{Single}}$ must be minimized. Note that each Client – Aide supplies lookup tables exclusively for their specific batch of training data.

Reducing the Size of Lookup Tables. We harness recent advancements in low precision machine learning [39, 71], discussed in Appendix B, to trim the table size. Neural networks trained with 16-bit fixed-point values retain their accuracy [39]. By adopting 16-bit values for activation function inputs and outputs, our lookup tables are constrained to 2^{16} entries.

Our experiments on the MNIST dataset (classifying as “0” or “not 0”) and the Arcene dataset (distinguishing “cancer” from “normal”) validate that 16-bit fixed-point numbers suffice for activation functions without compromising model accuracy. Figure 2 shows accuracy results for privacy-preserving logistic regression (PPLR) using our protocols, which employ 16-bit fixed-point representation for Sigmoid, and provides a comparison with plaintext training using 32-bit floating-point numbers. We observe that our protocol retains the model accuracy. Specifically, allocating 11 to the integer part for the fractional component and the remaining for the integer part yields accuracy nearly identical to plaintext training.

Optimizing Lookups for Derivatives. In certain cases, we can compute the derivative from the activation function or vice versa. For instance, $\text{DSigmoid}(x) = \text{Sigmoid}(x)(1 - \text{Sigmoid}(x))$ and $\text{ReLU}(x) = x\text{DReLU}(x)$, where DSigmoid and DReLU represent the derivative of Sigmoid and ReLU respectively. Our protocol, which leverages this relationship for ReLU and DReLU , is detailed in Algorithm 7. Once the parties have $\text{DReLU}(x)$, they can use a secure multiplication to compute $\text{ReLU}(x)$. This substantially cuts down the need for additional lookup tables. While the concept of computing $\text{ReLU}(x)$ from $\text{DReLU}(x)$ is not novel [76], our contribution lies in harnessing this relationship to halve the lookup tables needed for neural networks, thereby streamlining the offline phase.

An Alternate Representation for DReLU . We present an alternate representation of the DReLU activation, aiming to reduce the size of its associated lookup table substantially. Note that $\text{DReLU}(x) = 1$ if $x > 0$ and 0 otherwise. Our key observation is that, unlike other activation functions, DReLU does not require precise numbers. For instance, given $x_1 = 2$ and $x_2 = 2.57$, both yield $\text{DReLU}(x_1) = \text{DReLU}(x_2) = 1$. This suggests that the fractional component of the number can be disregarded without affecting the computation of DReLU . However, for $0 < x < 1$, eliminating the fractional component results in $x = 0$, leading to incorrect computation of $\text{DReLU}(x) = 0$. To address this, we propose the following representation for DReLU for fixed point numbers: $\text{DReLU}(x) = 1$ iff $\lfloor x + (2^{l_d} - 1) \rfloor \geq 1$ and 0 otherwise, where l_d represents bit count for the decimal part and $\lfloor \cdot \rfloor$ signifies truncation by l_d bits. This novel representation can substantially reduce the DReLU lookup table entries, allowing us to drop fractional bits and only retain integer bits. For instance, if l_d bits represent the fractional segment of a number and l_i bits correspond to the integer segment, a lookup table requires $2^{l_d+l_i}$ entries. With our proposed representation, only 2^{l_i} entries are needed, reducing the lookup table size for DReLU by a factor of 2^{l_d} . In our context, this factor is 2^{13} . As highlighted in previous work [58], each participant P_i can locally truncate their share of x . However, as this truncation introduces a small error (refer to Theorem 1 in [58] for details), we opt to retain one fractional bit instead of truncating all of them.

Addressing Storage Overhead. Despite our optimizations, the storage demands of lookup tables render $\text{HAWK}_{\text{Single}}$ suitable only for logistic regression and secure inference but not for neural network training, as shown in Table 5. To address this, we introduce $\text{HAWK}_{\text{Multi}}$, which permits multiple reuses of a single lookup table. While this approach introduces some leakage regarding access patterns, we prove the leakage is bounded by ϵ - $d_{\mathcal{X}}$ -privacy.

Algorithm 3 $\Pi_{\text{SC}(P_0, P_1)}^{\text{online}}$: Share Conversion Protocol

Input: For $i \in \{0, 1\}$, P_i knows n and N where $N > n$. P_i holds $\langle x \rangle_i^n$ and shares of a random r ; $\langle r \rangle_i^n$ and $\langle r \rangle_i^N$.

Output: P_i gets fresh share $\langle x \rangle_i^N$ of x .

- 1: P_i computes $\langle x \rangle_i^n - \langle r \rangle_i^n$ and sends it to P_{1-i}
 - 2: Each party reconstructs $z = x - r$.
 - 3: P_0 outputs $\langle x \rangle_0^N = z + \langle r \rangle_0^N$ and P_1 outputs $\langle x \rangle_1^N = \langle r \rangle_1^N$ as their share of x in \mathbb{Z}_N
-

Algorithm 4 Table Generation: $\Pi_{\text{HAWKMulti}}^{\text{offline}}(\text{Client-Aide})$

Input: A function $f : X \rightarrow Y$, total tables required m , cryptographic hash function H , field modulus n , order of an elliptic curve N , generator of the curve G .

Output: For $i \in \{0, 1\}$, P_i gets lookup tables $L_i^{f,c}$ for $0 \leq c < m$.

- 1 Client – Aide generates random keys k_i and $s_i \pmod N$ and sends to P_i .
- 2 For $0 \leq c \leq m - 1$:
 - 2.1 Client – Aide computes random $s_0^c = H(s_0 || c) \pmod N$.
 - 2.2 Client – Aide computes random $s_1^c = H(s_1 || c) \pmod N$.
 - 2.3 For all $x \in X$:
 - a Client – Aide computes $\kappa = k_0 k_1 \times (xG + s_0^c G + s_1^c G)$.
 - b Client – Aide computes additive secret shares $\langle y \rangle_0$ and $\langle y \rangle_1$ of $y = f(x)$.
 - c Client – Aide adds $\text{key} : H(\kappa)$, $\text{value} : \langle y \rangle_0$ in L_0^f .
 - d Client – Aide adds $\text{key} : H(\kappa)$, $\text{value} : \langle y \rangle_1$ in L_1^f .
- 2.4 Client – Aide randomly permutes all key-value pairs in $L_0^{f,c}$ and $L_1^{f,c}$.
- 3 Client – Aide sends all $L_0^{f,c}$ for $0 \leq c < m$ to P_0 .
- 4 Client – Aide sends all $L_1^{f,c}$ for $0 \leq c < m$ to P_1 .

Algorithm 5 Query Table: $\Pi_{\text{HAWKMulti}}^{\text{online}}(P_0, P_1)$

Input: For $i \in \{0, 1\}$, P_i holds random keys k_i and s_i , $\langle x \rangle_i^n$, a cryptographic hash function H , $\langle r \rangle_i^n$ and $\langle r \rangle_i^N$, order of an elliptic curve N , share of d_X -private noise $\langle y \rangle_i^n$.

Output: P_i gets secret share $\langle y \rangle_i$ of $y = f(x)$.

Initialization: P_i receives lookup tables $L_i^{f,c}$ for all $0 \leq c < m$ from Client – Aide, P_i initializes $c = 0$. P_i receives per query privacy budget ϵ and total privacy budget ϵ_T . P_i initializes $\epsilon_r = \epsilon_T$.

- 1: P_i computes noisy input $\langle \hat{x} \rangle_i^n = \langle x \rangle_i^n + \langle y \rangle_i^n$
- 2: P_i executes $\Pi_{\text{SC}(P_0, P_1)}^{\text{online}}$ with inputs $\langle \hat{x} \rangle_i^n$, $\langle r \rangle_i^n$ and $\langle r \rangle_i^N$ and receives output $\langle \hat{x} \rangle_i^N$.
- 3: P_0 computes random value $s_0^c = H(s_0 || c) \pmod N$.
- 4: P_1 computes random value $s_1^c = H(s_1 || c) \pmod N$.
- 5: P_0 computes $k_0 (\langle \hat{x} \rangle_0^N G + s_0^c G)$ and sends it to P_1 .
- 6: P_1 computes $k_1 (\langle \hat{x} \rangle_1^N G + s_1^c G)$ and sends it to P_0 .
- 7: P_0 computes $k_0 k_1 (\langle \hat{x} \rangle_1^N G + s_1^c G)$ and sends it to P_1 .
- 8: P_1 computes $k_1 k_0 (\langle \hat{x} \rangle_0^N G + s_0^c G)$ and sends it to P_0 .
- 9: P_i computes $\kappa = k_0 k_1 (\langle \hat{x} \rangle_0^N G + s_0^c G + s_1^c G) \equiv k_0 k_1 (\langle \hat{x} \rangle_1^N G + s_1^c G) + k_0 k_1 (\langle \hat{x} \rangle_0^N G + s_0^c G)$.
- 10: P_i retrieves the value: $\langle y \rangle_i$ corresponding to the key: $H(\kappa)$ from $L_i^{f,c}$, and outputs $\langle y \rangle_i$.
- 11: P_i computes $\epsilon_r = \epsilon_r - \epsilon$.
- 12: If $\epsilon_r == 0$:
 - 1.1 P_i increments $c = c + 1$.
 - 2.2 P_i sets $\epsilon_r = \epsilon_T$

5.2 HAWKMulti: Multi-Use Lookup Table Protocol

Algorithms 4 and 5 detail the offline and online phases of our HAWKMulti protocol, respectively. The core concept of the HAWKMulti protocol is the efficient reuse of a lookup table for multiple lookups. However, naively reusing a lookup table poses several challenges, as elaborated below.

Lookup Table Reuse. Directly reusing a lookup table to compute $f(x)$ for identical x values multiple times risks leaking the frequency distribution of x . While parties might not directly learn x , they can deduce a histogram representing each key's access frequency in the table. As demonstrated by Naveed et al. [60], this leakage may reveal plaintext when combined with auxiliary data.

Rather than entirely eliminating this access pattern leakage as in ORAM which has an inherent overhead of $\Omega(\log(n))$ [15, 36, 50], our HAWKMulti protocol employs d_X -privacy to limit it. This approach offers a balance between efficiency and privacy, precisely quantifying the leaked information. We analyze this leakage in detail in Section 6.1. With the HAWKMulti protocol, a lookup table remains reusable until the allocated privacy budget ϵ_T is depleted. Once exhausted, the table is discarded in favor of a new one.

Secure Deterministic Key Lookup. The objective is to securely compute a deterministic key PRF(K, x) from the shares $\langle x \rangle_i$ and use it to retrieve the value $\langle f(x) \rangle_i$ from the lookup table L_i^f . In the HAWKSingle, this is achieved by using the same pseudorandom pad $H(k_i || c)$ to obscure x for every key-value pair in a lookup table. This approach remains secure as each table is utilized once per activation function computation, and the pad is never reused.

In the HAWKMulti protocol, we utilize the hardness of the *elliptic curve discrete logarithm problem (ECDLP)* [43] to securely compute PRF(K, x). ECDLP, foundational to elliptic curve cryptography (ECC), posits that given elliptic curve points $P, Q \in E(\mathbb{F}_q)$, determining a such that $Q = aP$ is computationally challenging. ECC is integral to HAWKMulti for two additional properties, homomorphism and efficiency. ECC naturally supports homomorphic operations such as $xG = \langle x \rangle_0^N G + \langle x \rangle_1^N G$, where N denotes the elliptic curve order and G is the curve generator. We leverage this homomorphism to securely compute PRF(K, x). Furthermore, compared to RSA, ECC requires $12\times$ smaller keys for 128 bits of security, leading to faster computations and reduced communication [54].

We iteratively detail the protocol design, starting with the core concept and refining it to the finalized protocol. Suppose parties P_0 and P_1 possess shares $\langle x \rangle_0^N$ and $\langle x \rangle_1^N$ of x . They also have secret random keys k_0 and k_1 , respectively. In the subsequent section, we present a protocol to convert a share $\langle x \rangle_i$ of x from $\langle x \rangle_i^N$ to $\langle x \rangle_i^N$. P_0 computes $k_0 (\langle x \rangle_0^N G)$ and forwards it to P_1 which then computes $k_1 k_0 (\langle x \rangle_0^N G)$. Similarly, P_1 sends $k_1 (\langle x \rangle_1^N G)$ to P_0 which computes $k_0 k_1 (\langle x \rangle_1^N G)$. Due to the ECDL assumption, $k_i (\langle x \rangle_i^N G)$ does not leak information about $\langle x \rangle_i^N$ or k_i . Subsequently, P_1 sends $k_1 k_0 (\langle x \rangle_0^N G)$ to P_0 and P_0 sends $k_0 k_1 (\langle x \rangle_1^N G)$ to P_1 . Each party P_i can then compute $\kappa = k_0 k_1 (xG) \equiv k_0 k_1 (\langle x \rangle_1^N G) + k_0 k_1 (\langle x \rangle_0^N G)$. This allows P_0 and P_1 to securely compute PRF(K, x), where $K = k_0 k_1$ which can be used as a key for table lookup.

Converting Additive Shares To A Larger Group. For the HAWKMulti protocol, we start by converting additive shares from \mathbb{Z}_n to \mathbb{Z}_N , where $N > n$ and N is the order of the elliptic curve E . Algorithm 3 presents a one-round protocol for share conversion based on [25]. The protocol requires shares of a random r in both \mathbb{Z}_n and \mathbb{Z}_N , which can be precomputed in the offline phase.

Mitigating Malleability. Given $k_0 k_1 (xG)$, it is possible to compute $k_0 k_1 \cdot (axG) \equiv a \times (k_0 k_1 (xG))$ for any arbitrary a . While this

does not directly reveal x , it might leak crucial information in our protocol when multiple lookups are performed on a table. A malicious party could compute $k_0k_1(axG)$ for various a values and cross-check all past and future lookups $k_0k_1(bG)$ to determine if any $k_0k_1(axG) = k_0k_1(bG)$, thereby inferring if b is a multiple of x . To counter this, we use $k_0k_1(xG + s_0G + s_1G)$ as the lookup table key. Here k_0 and s_0 are random keys known only to P_0 , while k_1 and s_1 are known only to P_1 . Having these two sets of keys ensures no party P_i can independently compute $k_0k_1(axG + s_0G + s_1G)$. Furthermore, instead of directly using s_0 and s_1 , the parties use distinct s_0^c and s_1^c for each table where c denotes the lookup table index and $s_i^c = H(s_i || c)$. This guarantees that the lookup key differs for each x across different lookup tables.

Reducing Number of Elliptic Curve Scalar Multiplications.

Optimization 1. During the offline phase, a Client – Aide must generate m tables, each of size 2^{16} . A significant computational bottleneck in our $\text{HAWK}_{\text{Multi}}$ protocol arises when computing elliptic curve scalar multiplications in step 2.3a of Algorithm 4. This step requires four scalar multiplications and two point additions on the elliptic curve. This translates to a daunting $m \times 2^{18}$ scalar multiplications and $m \times 2^{17}$ point additions in total. We introduce an optimization to reduce these numbers drastically: scalar multiplications are cut down to $m + 1$, and point additions to $m \times 2^{16}$. Specifically, the Client – Aide initially calculates $(k_0k_1) \cdot G$. For a table $L^{f,c}$, it then computes $(k_0k_1)(s_0^c + s_1^c) \cdot G$. The key for $x = 0$ is set as $\text{PRF}(K, 0) = (k_0k_1)(s_0^c + s_1^c) \cdot G$. All subsequent keys are iteratively determined as $\text{PRF}(K, i) = \text{PRF}(K, i - 1) + (k_0k_1) \cdot G$, requiring only a single point addition.

Optimization 2. In Algorithm 5, steps 5 and 6, if executed directly, demand three scalar multiplications and one point addition for each party on the elliptic curve. We suggest an optimization to bring this down to a single scalar multiplication per party. Each party P_i first evaluates $k_i(\langle x \rangle_i^N + s_i^c)$, and subsequently computes $k_i(\langle x \rangle_i^N + s_i^c) \cdot G$, which requires just one scalar multiplication.

Ensuring Differentially Private Access Patterns. While our $\text{HAWK}_{\text{Multi}}$ protocol conceals the input x during lookups, repeated lookups on the same table might inadvertently reveal a histogram of the frequency of each pseudorandom key's access in the table. To address this, our $\text{HAWK}_{\text{Multi}}$ protocol uses $d_{\mathcal{X}}$ -privacy to limit this leakage. Rather than using their actual input share $\langle x \rangle_i$, each party computes a perturbed input share $\langle \hat{x} \rangle_i$ by incorporating $d_{\mathcal{X}}$ -private noise $\langle \gamma \rangle_i$ as outlined in step 1 of Algorithm 5. This ensures that the parties only learn noisy access patterns, maintaining ϵ - $d_{\mathcal{X}}$ -privacy. The shares of noise γ are generated in the offline phase.

5.3 Computing Activation Functions

We now describe how we compute different activation functions using our protocols. We use $\Pi_{\text{Lookup}}^{\text{online}}$ to denote either $\Pi_{\text{HAWK}_{\text{Single}}}^{\text{online}}$ or $\Pi_{\text{HAWK}_{\text{Multi}}}^{\text{online}}$.

ReLU and DReLU. Algorithm 6 and 7 describe our protocols for computing DReLU and ReLU respectively. We use the property: $\text{ReLU}(x) = x\text{DReLU}(x)$, to compute $\text{ReLU}(x)$ from $\text{DReLU}(x)$.

Algorithm 6 DReLU: $\Pi_{\text{DReLU}}(P_0, P_1)$

Input: P_i holds $\langle x \rangle_i$, lookup table L_i^{DReLU} .

Output: P_i gets $\langle z \rangle_i = \langle \text{DReLU}(x) \rangle_i$

- 1: P_i calls $\Pi_{\text{Lookup}}^{\text{online}}(P_i, P_{1-i})$ with inputs $(\langle x \rangle_i + i \times (2^{1d} - 1), L_i^{\text{DReLU}})$
 - 2: P_i learns $\langle z \rangle_i = \langle \text{DReLU}(x) \rangle_i$.
-

Algorithm 7 ReLU: $\Pi_{\text{ReLU}}(P_0, P_1)$

Input: P_i holds $\langle x \rangle_i$, $\langle \text{DReLU}(x) \rangle_i$.

Output: P_i gets $\langle z \rangle_i = \langle \text{ReLU}(x) \rangle_i$

- 1: For $i \in \{0, 1\}$, P_i calls $\Pi_{\text{SecMul}}(P_i, P_{1-i})$ having inputs $(\langle \text{DReLU}(x) \rangle_i, \langle x \rangle_i)$ and learns $\langle z \rangle_i = \langle x\text{DReLU}(x) \rangle_i$. Here Π_{SecMul} denotes secure multiplication using Beaver's triples.
 - 2: P_i outputs $\langle z \rangle_i = \langle \text{ReLU}(x) \rangle_i = \langle x\text{DReLU}(x) \rangle_i$.
-

Algorithm 8 Sigmoid: $\Pi_{\text{Sigmoid}}(P_0, P_1)$

Input: P_i holds $\langle x \rangle_i$, lookup table L_i^{Sigmoid} .

Output: P_i gets $\langle z \rangle_i = \langle \text{Sigmoid}(x) \rangle_i$.

- 1: For $i \in \{0, 1\}$, P_i calls $\Pi_{\text{Lookup}}^{\text{online}}(P_i, P_{1-i})$ with inputs $(\langle x \rangle_i, L_i^{\text{Sigmoid}})$
 - 2: P_i learns $\langle z \rangle_i = \langle \text{Sigmoid}(x) \rangle_i$.
-

Algorithm 9 Softmax: $\Pi_{\text{Softmax}}(P_0, P_1)$

Input: For $i \in \{0, 1\}$ and $j \in \{1, \dots, d_I\}$, P_i holds $\langle x_j \rangle_i$, lookup tables L_i^{exp} and L_i^{inverse} .

Output: For neuron $j \in \{1, \dots, d_I\}$, P_i gets $\langle \text{Softmax}(x_j) \rangle_i$.

- 1: For $j \in \{1, \dots, d_I\}$, P_i calls $\Pi_{\text{Lookup}}^{\text{online}}(P_i, P_{1-i})$ with inputs $(\langle x_j \rangle_i, L_i^{\text{exp}})$ and learns $\langle e^{x_j} \rangle_i$.
 - 2: P_i locally computes $\langle s \rangle_i = \sum_{j=1}^{d_I} \langle e^{x_j} \rangle_i$ as the sum of the exponential terms.
 - 3: P_i calls $\Pi_{\text{Lookup}}^{\text{online}}(P_i, P_{1-i})$ with inputs $(\langle s \rangle_i, L_i^{\text{inverse}})$ and learns $\langle \frac{1}{s} \rangle_i$ as the share of the inverse of the sum s .
 - 4: For $j \in \{1, \dots, d_I\}$, P_i runs $\Pi_{\text{SecMul}}(P_i, P_{1-i})$ with inputs $(\langle e^{x_j} \rangle_i, \langle \frac{1}{s} \rangle_i)$ and learns $\langle \text{Softmax}(x_j) \rangle_i$.
-

Sigmoid. Algorithm 8 describes our protocol to compute $\text{Sigmoid}(x)$. It is worth noting that the prior works [58, 76] do not provide any method to compute Sigmoid accurately.

Softmax. Algorithm 9 describes our protocol for computing Softmax. As Softmax involves d_I inputs, using our protocols naively requires $2^{n \times d_I}$ entries in the lookup table given x_j is an n -bit number, which is impractical. We introduce a method that substantially reduces the lookup table size, albeit with a slight increase in online computation. Rather than directly computing Softmax using a single lookup table, we employ two distinct tables: one for the exponential function, $\exp(x) = e^x$, and another for the inverse, $\text{inverse}(x) = \frac{1}{x}$. Existing literature lacks a precise method for Softmax computation, resorting to non-standard functions [58, 76], which leads to accuracy degradation (see Figure 9).

Protocol	Rounds		Communication	
	HAWK _{Single}	HAWK _{Multi}	HAWK _{Single}	HAWK _{Multi}
Π_{Lookup}	1	3	l	$l + 2l'$
Π_{DReLU}	1	3	l	$l + 2l'$
Π_{ReLU} (after DReLU)	1	1	l	l
Π_{Sigmoid}	1	3	l	$l + 2l'$
Π_{Softmax}	3	7	$l(1 + 2d_l)$	$l(1 + 2d_l) + 2l'(1 + d_l)$

Table 1: Total rounds and communication complexity of HAWK_{Single} and HAWK_{Multi} protocols. ($l = 64$ bits, $l' = 256$ bits)

5.4 Rounds and Communication Complexity

Table 1 details the rounds and communication complexities of sub-protocols employed in training both logistic regression models and neural networks. For the HAWK_{Single} protocol, with l -bit numbers ($l = 64$), the associated protocols: Π_{Lookup} , Π_{Sigmoid} and, Π_{DReLU} , demand a single communication round with a complexity of l bits. The Π_{ReLU} requires an extra round due to an additional secure multiplication following DReLU. Meanwhile, Π_{Softmax} requires three rounds, attributed to the computations of the exponential function, the inverse function, and a secure multiplication. Here, d_l denotes the number of neurons in the output layer.

For the HAWK_{Multi} protocol, we adopt numbers of $l' = 256$ bits. The associated protocols: Π_{Lookup} , Π_{Sigmoid} , and Π_{DReLU} , require three communication rounds with a complexity of $l + 2l'$. The HAWK_{Multi} based protocol Π_{Softmax} requires a total of seven rounds.

The efficiency improvements in training PPML models in our case, relative to SecureML, stem from the fast computation of non-linear activations such as ReLU, Sigmoid, Softmax, and their derivatives. Past works [46, 53, 58] leveraged Yao's garbled circuits for these computations. Transitioning secret shares between arithmetic and Yao sharing introduces an overhead of $7l\kappa + (l^2 + l)/2$ [26].

6 SECURITY ANALYSIS

Our privacy-preserving training and inference protocols primarily modify SecureML's activation function computation, replacing it with our HAWK_{Single} or HAWK_{Multi} based protocols: Π_{Sigmoid} , Π_{ReLU} , Π_{DReLU} , and Π_{Softmax} . Each protocol either exclusively employs $\Pi_{\text{HAWKSingle}}^{\text{online}}$ or $\Pi_{\text{HAWKMulti}}^{\text{online}}$, or combines them with secure multiplication. Consequently, our security analysis focuses on the $\Pi_{\text{HAWKSingle}}^{\text{online}}$ and $\Pi_{\text{HAWKMulti}}^{\text{online}}$ protocols, given their composability. The overall security of our protocols follows from SecureML's subprotocols composed with our $\Pi_{\text{HAWKSingle}}^{\text{online}}$ or $\Pi_{\text{HAWKMulti}}^{\text{online}}$ protocols.

Functionality $\mathcal{F}_{\text{SingleLookup}}$
Parameters: Servers P_0, P_1
Data: On input $\langle x \rangle_0$ and $\langle x \rangle_1$, store $\langle x \rangle_0$ and $\langle x \rangle_1$ internally.
Computation: On input f from P_0 or P_1 , reconstruct $x = \langle x \rangle_0 + \langle x \rangle_1 \pmod{n}$, compute $y = f(x)$, create additive shares $\langle y \rangle_0$ and $\langle y \rangle_1$, and send $\langle y \rangle_0$ to P_0 and $\langle y \rangle_1$ to P_1 .

Figure 3: Ideal Functionality $\mathcal{F}_{\text{SingleLookup}}$

Functionality \mathcal{F}_{SC}

Parameters: Servers P_0, P_1

Data: On input $\langle x \rangle_0^n$ and $\langle x \rangle_1^n$, store $\langle x \rangle_0^n$ and $\langle x \rangle_1^n$ internally.

Computation: On input $N > n$ from P_0 or P_1 , reconstruct $x = \langle x \rangle_0^n + \langle x \rangle_1^n \pmod{n}$, create random additive shares $\langle x \rangle_0^N$ and $\langle x \rangle_1^N$ of x , and send $\langle x \rangle_0^N$ to P_0 and $\langle x \rangle_1^N$ to P_1 .

Figure 4: Ideal Functionality \mathcal{F}_{SC}

Functionality $\mathcal{F}_{\text{MultiLookup}}$

Parameters: Servers P_0, P_1

Data: On inputs $\langle x \rangle_0$ and $\langle x \rangle_1$, store internally.

Computation:

- 1 Reconstruct $x = \langle x \rangle_0 + \langle x \rangle_1 \pmod{n}$.
- 2 Compute $\hat{x} = x + \gamma \pmod{n}$, with γ drawn using geometric mechanism in Definition 4.
- 3 On input f from P_0 or P_1 , compute $y = f(\hat{x})$.
- 4 Create additive shares $\langle y \rangle_0$ and $\langle y \rangle_1$ of y .
- 5 Send $\langle y \rangle_0$ to P_0 and $\langle y \rangle_1$ to P_1 .

Figure 5: Ideal Functionality $\mathcal{F}_{\text{MultiLookup}}$

THEOREM 1. Protocol $\Pi_{\text{HAWKSingle}}^{\text{online}}$ in Algorithm 2 securely realizes the functionality $\mathcal{F}_{\text{SingleLookup}}$ against semi-honest adversaries.

PROOF SKETCH. Consider an adversary \mathcal{A} that corrupts either P_0 or P_1 . Without loss of generality, let's focus on the case where \mathcal{A} corrupts P_0 . We construct a simulator \mathcal{S} that simulates \mathcal{A} in the ideal world. The simulator \mathcal{S} runs \mathcal{A} . \mathcal{S} generates a random value r to mimic the honest server P_1 's message $\langle x \rangle_1 + H(k_1 || c) \pmod{n}$ and forwards it to \mathcal{A} . To simulate the honest server's output, \mathcal{S} uses $\langle y \rangle_1$ received from $\mathcal{F}_{\text{SingleLookup}}$.

The indistinguishability between the real and ideal worlds stems from the fact that the sole message exchanged, $\langle x \rangle_1 + H(k_1 || c) \pmod{n}$, is computationally indistinguishable from a random value r , owing to the one-time pseudorandom pad $H(k_1 || c)$. \square

THEOREM 2. Protocol $\Pi_{\text{SC}}^{\text{online}}$ in Algorithm 3 securely realizes the functionality \mathcal{F}_{SC} against semi-honest adversaries.

PROOF SKETCH. Given the symmetry of our protocol with respect to both servers, we focus on the scenario where an adversary corrupts party P_0 . To simulate the honest server P_1 in the ideal world, the simulator \mathcal{S} generates a random value s representing the message $\langle x \rangle_1^n + \langle r \rangle_1^n \pmod{n}$ and sends it to \mathcal{A} . \mathcal{S} employs $\langle x \rangle_1^N$ obtained from \mathcal{F}_{SC} to simulate the honest server's output.

The indistinguishability between the real and ideal worlds stems from the fact that the sole message exchanged is $\langle x \rangle_1^n + \langle r \rangle_1^n \pmod{n}$, with $\langle r \rangle_1^n$ being a random share of r . Consequently, $\langle x \rangle_1^n + \langle r \rangle_1^n \pmod{n}$ is computationally indistinguishable from a random s of equivalent length. \square

6.1 Leakage Analysis

We begin by discussing the leakage function \mathcal{L} for our Hawk_{Multi} protocol. Intuitively, our Hawk_{Multi} protocol leaks a perturbed profile of access patterns to a lookup table $L^{f,c}$, where c is the table index. Note that the protocol does not reveal the input x to any party. Instead, it only reveals a distorted frequency distribution of accesses to a table $L^{f,c}$ without associating access to a specific input x due to the pseudorandom nature of the key κ . For instance, if we access a lookup table $r_{\text{Multi}} = 10$ times, each with ϵ - d_X -privacy, before exhausting our total privacy budget ϵ_T , the Hawk_{Multi} protocol merely discloses if the ten lookups were distinct or if there were repeated lookups for a particular key κ . Furthermore, adding noise γ to x before each lookup ensures that parties remain unaware if a key repetition was genuine or noise-induced. We employ the geometric mechanism in Definition 4 to sample the noise γ .

DEFINITION 7. *The leakage function \mathcal{L} is κ , where κ is a pseudo-random key for a lookup table $L^{f,c}$ and $\kappa = k_0 k_1 (\hat{x}G + s_0^c G + s_1^c G)$, where k_i and s_i^c are secret random keys of a party P_i , G is the generator of an elliptic curve E , and $\hat{x} = x + \gamma$, with γ being random noise as per Definition 4.*

THEOREM 3. *The leakage function \mathcal{L} is local ϵ - d_X -private as per Definition 3.*

The preservation of local ϵ - d_X -privacy by \mathcal{L} is a consequence of our utilization of the geometric mechanism, as defined in Definition 4, to sample noise γ . We prove this mechanism to preserve ϵ - d_X -privacy in Appendix C. Given the post-processing resilience of differential privacy [31, 32], \mathcal{L} also maintains ϵ - d_X -privacy.

THEOREM 4. *Protocol $\Pi_{\text{HawkMulti}}^{\text{online}}$ in Algorithm 5 securely realizes the functionality $\mathcal{F}_{\text{MultiLookup}}$ with \mathcal{L} leakage in $(\mathcal{F}_{\text{SC}})$ -hybrid model as per Definition 6.*

PROOF SKETCH. Consider an adversary \mathcal{A} that corrupts either server P_0 or P_1 . Given the protocol’s symmetry concerning both servers, we focus on the scenario where \mathcal{A} corrupts P_0 . We describe a simulator \mathcal{S} to simulate \mathcal{A} in the ideal world. The simulator \mathcal{S} runs \mathcal{A} . \mathcal{S} employs random s to simulate the honest server P_1 ’s output of \mathcal{F}_{SC} . \mathcal{S} also generates random points R and S on the elliptic curve E to simulate P_1 ’s messages $k_1(\langle \hat{x} \rangle_1^N G + s_1^c G)$ and $k_1 k_0(\langle \hat{x} \rangle_0^N G + s_0^c G)$ and sends it to \mathcal{A} . \mathcal{S} also sends leakage function \mathcal{L} in Definition 7 to \mathcal{A} . \mathcal{S} uses $\langle y \rangle_1$ received from $\mathcal{F}_{\text{MultiLookup}}$ to simulate the honest server’s output.

The adversary \mathcal{A} ’s view in the ideal and real worlds is indistinguishable because $\langle x \rangle_1^N$ is computationally indistinguishable from a random value s . Moreover, according to the *elliptic curve discrete log assumption* $k_1(\langle \hat{x} \rangle_1^N G + s_1^c G)$ and $k_1 k_0(\langle \hat{x} \rangle_0^N G + s_0^c G)$ are indistinguishable from random points R and S on E as $\langle \hat{x} \rangle_0^N, \langle \hat{x} \rangle_1^N, k_0, k_1, s_0^c$, and s_1^c are uniform random in \mathbb{Z}_N . The only leakage in the real world and the ideal world is \mathcal{L} as defined in Definition 7 which is proved to preserve ϵ - d_X -privacy. \square

7 END-TO-END TRAINING PROTOCOLS

7.1 Privacy-Preserving Logistic Regression

Logistic regression necessitates the computation of the Sigmoid function. While previous studies have approximated Sigmoid using

polynomials [4], achieving satisfactory accuracy demands high-degree polynomials, leading to inefficiencies. SecureML uses a piecewise function and Yao’s garbled circuits for Sigmoid approximation, but the overhead and accuracy loss from this method remain a concern. We introduce the Π_{Sigmoid} protocol to directly compute Sigmoid, enabling efficient training of logistic regression models without compromising plaintext training accuracy. Algorithm 10 in Appendix D presents our complete PPLR protocol. After the dot product computation of input examples with coefficients during the forward pass, parties employ Π_{Sigmoid} for Sigmoid computation, subsequently proceeding to backward propagation.

7.2 Privacy-Preserving Neural Network

Training neural networks securely introduces complexity and inefficiency primarily during the computation of non-linear activations, such as ReLU, DReLU, and Softmax. SecureML employs Yao’s garbled circuits for ReLU and DReLU computations. They also suggest an approximation for Softmax as $f(z_i) = \frac{\text{ReLU}(z_i)}{\sum_{j=1}^{d_I} \text{ReLU}(z_j)}$. However, this function demands a division garbled circuit, leading to inefficiencies, and the approximation compromises accuracy, as highlighted in Section 8.4.

In Section 5.3, we introduce efficient protocols for computing ReLU, DReLU, and Softmax. We can adeptly train diverse neural networks by integrating the methods from Section 7.1 and the protocols in Section 5.3. For instance, SecureML’s 3-layer neural network, comprising two fully connected layers with ReLU activation and an output layer with Softmax activation, can be implemented in the forward pass using Π_{SecMul} , Π_{ReLU} , and Π_{Softmax} . Back-propagation, on the other hand, requires Π_{SecMul} and Π_{DReLU} . Although our framework is readily extendable to CNNs, we focus on DNNs in our evaluation and defer CNNs to future work.

8 EVALUATION

8.1 Implementation

We implement our system in C++ with $\sim 13,570$ lines of code. We use the Eigen library [38] for matrix operations, libsecp256k1 [82] for EC implementation, and Crypto++ [23] for cryptographic operations.

We adopt a field size of 2^{64} and represent numbers using C++’s native unsigned long integer datatype. Following SecureML’s fixed-point representation [58], we allocate 13 bits to the fractional component. For lookup tables, numbers are represented with 16 bits divided between integer and fractional parts. The lookup tables for Sigmoid allocate 3 bits for the integer and 13 for the fractional part. For DReLU tables, based on the alternate representation from Section 5.1, we use 4 integer bits and 1 fractional bit. For the Softmax tables, the exponential function is represented with 6 integer and 10 fractional bits, while the inverse function uses 14 integer and 2 fractional bits, ensuring the necessary high dynamic range. In our experiments involving Hawk_{Multi} protocol, N is set to match the order of the SECP256K1 curve.

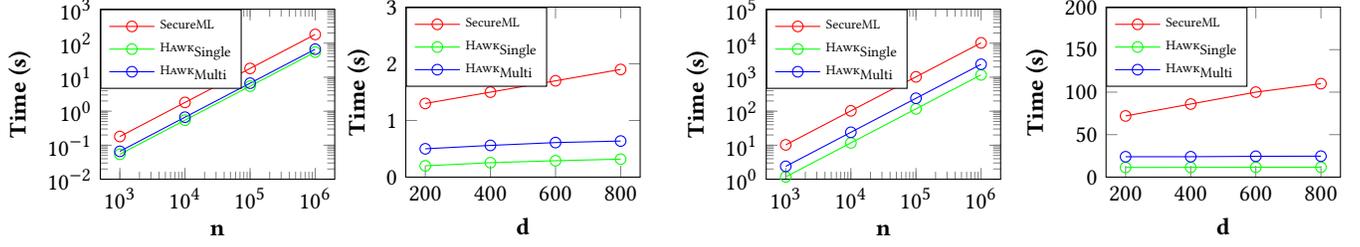


Figure 6: a) LAN setting (d=784) b) LAN setting (n=10,000) c) WAN setting (d=784) d) WAN setting (n=10,000) Comparison of our PPLR protocols with SecureML. Here n denotes number of data records, and d denotes the data dimension.

8.2 Experimental Setup

Our experiments utilize two “Amazon EC2 c4.8x large” instances, aligning with the setup in previous work [58]. We assess performance in both LAN and WAN settings:

LAN setting: Both “Amazon EC2 c4.8x large” instances are hosted within the same region, achieving an average bandwidth of 620 MB/s and a network delay of 0.25ms.

WAN setting: The instances are hosted in separate regions: US East and US West. This configuration yields an average bandwidth of 32 MB/s and a network delay of 48 ms.

We use these datasets to evaluate our protocols: MNIST [51], Gisette [41, 42], Arcene [40, 42], Fashion MNIST [83], Iris [33], and the Adult dataset [9]. We describe each dataset in Appendix E.

Each data point represents the average of 10 runs. To assess scalability, we create synthetic datasets from MNIST by altering record count and data dimensions. For our HAWKMulti protocol, we experiment by adjusting the total privacy budget, ϵ_T , of the lookup tables, where $\epsilon_T = \epsilon \times r_{Multi}$, with ϵ as the per-access privacy budget and r_{Multi} denoting each lookup table’s reuse frequency.

8.3 Privacy Preserving Logistic Regression

This section evaluates our PPLR protocol’s performance and scalability, compares it with SecureML [58], and assesses the accuracy against plaintext training.

Framework	MNIST	Gisette	Arcene	Fashion MNIST
Tensorflow	99.19	98	86	95.97
HAWKSingle	99.21	98.1	86	95.97
HAWKMulti ($\epsilon_T = 0.01$)	99.21	98.1	86	95.97
HAWKMulti ($\epsilon_T = 0.001$)	99.20	97.1	85	95.89
HAWKMulti ($\epsilon_T = 0.0005$)	98.80	95.2	85	95.51

Table 2: Accuracy comparison of our PPLR protocols with plaintext training ($r_{Multi} = 10$).

Accuracy Performance. Table 2 indicates our HAWKSingle protocol matches plaintext training accuracy across various datasets. For HAWKMulti, there is no accuracy loss at $\epsilon_T = 0.01$ and only a minor decline at $\epsilon_T = 0.0005$, suggesting that HAWKMulti offers robust privacy without significant accuracy compromises. Figure 7 illustrates the training trajectory on MNIST and Gisette datasets

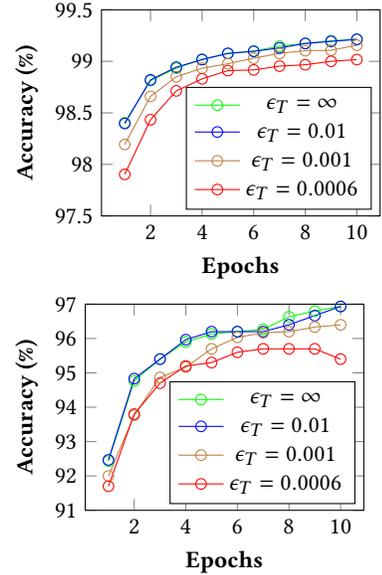


Figure 7: a) MNIST dataset b) Gisette dataset Training trajectory for PPLR across various ϵ_T values

Framework	LAN (s)	WAN (s)	Offline (s)	Storage
SecureML	10.84	612.78	-	-
HAWKSingle	3.2	70.2	57.58	36.62 GB
HAWKMulti ($\epsilon_T = 0.01$)	4	144	18.3	0.37 GB

Table 3: Comparison of PPLR with SecureML on MNIST ($|B| = 128$, $r_{Multi} = 100$). SecureML does not report the offline cost.

with HAWKMulti, highlighting the influence of varying ϵ_T . Comparing with SecureML [58], in the task of classifying MNIST digits as zero vs. non-zero, our protocols achieve 99.21% accuracy, closely followed by a Tensorflow model at 99.19%. In contrast, SecureML reaches 98.62%. Notably, SecureML’s accuracy suffers more from non-standard activations, especially with neural networks, with a drop of ~ 3% for the MNIST dataset, as discussed in Section 8.4.

Computation and Communication Performance. Table 3 details our PPLR experiments in LAN and WAN environments. The online phase for HAWKSingle requires 3.2s (LAN) and 70.2s (WAN)

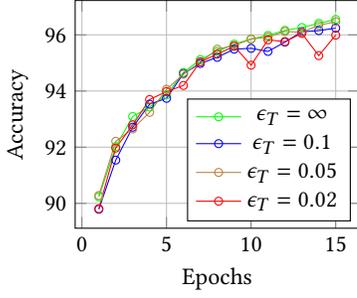


Figure 8: Training trajectory for PPNN across various ϵ_T values for MNIST.

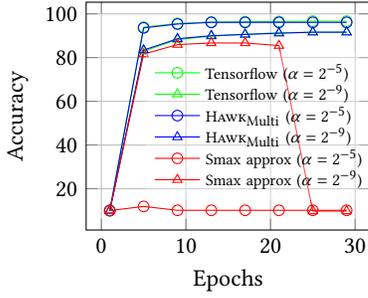


Figure 9: Accuracy comparison of PPNN training on MNIST. Here α denotes learning rate and for $\text{HAWK}_{\text{Multi}}$, $\epsilon_T = 0.1$. The Softmax approximation used by past works causes accuracy drop.

with a lookup table storage overhead of 36.62GB per party. In contrast, $\text{HAWK}_{\text{Multi}}$ has an online duration of 4s (LAN) and 144s (WAN) but reduces storage cost to 0.37GB for $r_{\text{Multi}} = 100$. This table underlines the balance between privacy, offline cost, and storage. While $\text{HAWK}_{\text{Single}}$ offers perfect security with a faster online phase, $\text{HAWK}_{\text{Multi}}$ demands fewer lookup tables, ensuring $d_{\mathcal{X}}$ -privacy against access pattern leaks. Relative to SecureML, our methods are up to 3.4x and 8.7x faster in LAN and WAN, respectively.

For scalability experiments, we vary the dataset size n and the data dimension d . Figure 6 illustrates that PPLR training time scales linearly with n . As d grows, our PPLR time increases sublinearly, while SecureML’s rises linearly. In WAN, with d ranging from 200 to 1000, SecureML’s time increases from 72s to 124s. In contrast, our protocols remain steady at around 12s for $\text{HAWK}_{\text{Single}}$ and 24s for $\text{HAWK}_{\text{Multi}}$, as d increases. In WAN, the overall time is largely influenced by communication time, which, for our protocols, stays nearly constant as d increases.

8.4 Privacy Preserving Neural Networks

To evaluate our protocols for training neural networks, we consider an architecture similar to prior works [58, 76]. The neural network is fully connected, consisting of 2 hidden layers, each with 128 neurons and an output layer with 10 neurons. The hidden layers use ReLU, and the output layer uses Softmax activation.

Framework	MNIST	Iris	Adult
Tensorflow	96.6	96.56	85.4
$\text{HAWK}_{\text{Single}}$	96.6	95.56	85.4
$\text{HAWK}_{\text{Multi}}$ ($\epsilon_T = 0.1$)	96.34	96.56	85.4
$\text{HAWK}_{\text{Multi}}$ ($\epsilon_T = 0.05$)	96.09	96.56	85.2
$\text{HAWK}_{\text{Multi}}$ ($\epsilon_T = 0.02$)	95.99	96.56	85.2

Table 4: Accuracy comparison of our PPNN protocols with plaintext training ($r_{\text{Multi}} = 10$).

Accuracy Performance. The neural network trained on MNIST using our PPNN protocol can reach an accuracy of 96.6% within 15 epochs, similar to a Tensorflow model. Neural networks trained on Iris [33] and the Adult dataset [9] using our $\text{HAWK}_{\text{Single}}$ protocol closely match plaintext training accuracies, as shown in Table 4. However, with $\text{HAWK}_{\text{Multi}}$, accuracy slightly diminishes as ϵ_T decreases. For instance, at $\epsilon_T = 0.02$, MNIST accuracy drops by $\sim 0.5\%$. Figure 8 depicts the MNIST training trajectory under $\text{HAWK}_{\text{Multi}}$, emphasizing the effect of different ϵ_T values. Additionally, Figure 11 illustrates the correlation between the total number of lookup tables required and various privacy budget thresholds necessary to achieve specific accuracy levels. As expected $\text{HAWK}_{\text{Single}}$ protocol requires the highest number of lookup tables. For the $\text{HAWK}_{\text{Multi}}$ protocol, a decrease in ϵ_T in case of the Adult dataset necessitates an increase in the number of lookup tables to maintain the same level of accuracy.

Previous studies [57, 58, 76] report around 93.4% accuracy on MNIST using the same setup. This discrepancy primarily stems from their use of a non-standard activation function for Softmax. Figure 9 contrasts our training protocol’s accuracy with plaintext training. When we emulate the Softmax approximation ($f(z_i) = \frac{\text{ReLU}(z_i)}{\sum_{j=1}^d \text{ReLU}(z_j)}$) from prior works [57, 58, 76] in a plaintext NN model, we notice a decline in accuracy and delayed convergence. Notably, our protocol surpasses 90% accuracy on MNIST in under 1.06 epochs, while earlier works [57, 58, 76] need more epochs for comparable results. As shown in Figure 9, our protocol benefits from a larger learning rate, leading to quicker convergence. In contrast, the Softmax approximation can cause accuracy to plummet after a certain number of epochs. This decline, consistent across various learning rates, suggests that such approximations might be even more detrimental for more complex tasks, potentially causing substantial accuracy losses.

Computation and Communication Performance. Table 5 compares our PPNN protocols with SecureML [58]. While we also reference other works [1, 57, 76, 77] that encompass both 2PC and 3PC protocols, a direct comparison is challenging due to differing settings, as highlighted in Section 2. In LAN, our PPNN protocol’s online phase outperforms SecureML’s client-aided approach by 26 – 144x. SecureML, in the WAN context, omits client-aided protocol training times, offering only metrics for the setting with server-generated random shares. This method increases offline costs but reduces online time through matrix multiplication speed gains. Our WAN-based PPNN protocol is 159 – 281x faster than SecureML’s server-aided variant. For the client-aided variant, factoring in a 2.45 slowdown from the LAN setting, our protocol’s speed advantage

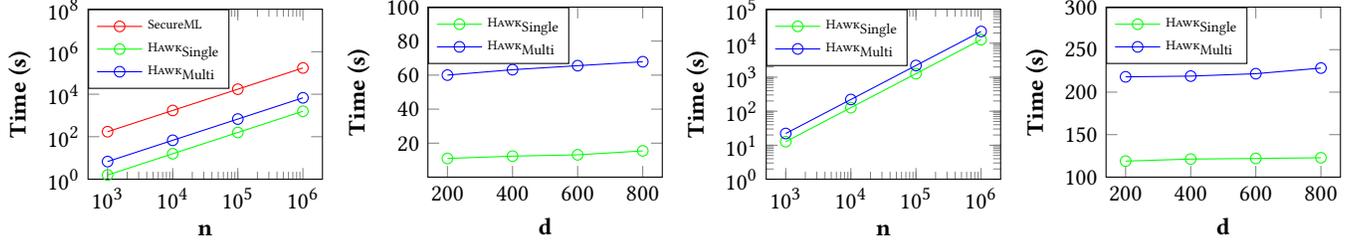


Figure 10: a) LAN setting ($d=784$) b) LAN setting ($n=10,000$) c) WAN setting ($d=784$) d) WAN setting ($n=10,000$)
 Comparison of our PPNN protocols with SecureML. Here, n denotes number of data records, and d denotes the data dimension.

Framework	LAN		WAN		Lookup Table Storage
	Offline (hours)	Online (hours)	Offline (hours)	Online (hours)	
HAWK _{Single}	0.47	0.02	0.47	0.21	407.5 GB
HAWK _{Multi} ($\epsilon_T = 1$)	0.35	0.11	0.35	0.37	4.1 GB
2PC SecureML (server-aided) [58]	80.5	1.17	4277	59	-
SecureML (client-aided) [58]	4.15	2.87	-	-	-
Quotient [1]	-	8.72	-	50.74	-
3PC Falcon [77]	-	0.17	-	3.76	-
ABY ³ [57]	-	0.75	-	-	-
SecureNN [76]	-	1.03	-	7.83	-

Table 5: Comparison of privacy-preserving neural network training using MNIST dataset ($|\mathcal{B}| = 128$, $r_{\text{Multi}} = 100$).

Framework	Logistic Regression				Neural Networks			
	LAN (ms)		WAN (ms)		LAN (s)		WAN (s)	
	Offline	Online	Offline	Online	Offline	Online	Offline	Online
SecureML (server-aided)	51	3.9	2010	429	13.8	0.20	472	1.2
HAWK _{Single}	96	3.7	96	171	1.43	0.08	1.43	0.82

Table 6: Secure inference time comparison with SecureML on MNIST for 100 samples.

over SecureML in WAN surges to $391 - 688\times$. While our PPNN protocol drastically reduces offline costs compared to SecureML, our HAWK_{Single} adds a lookup table storage overhead of 407.5GB. In contrast, our HAWK_{Multi} reduces this overhead to 4.1GB per party. Compared with Quotient [1], another 2PC protocol, our PPNN protocol’s online phase is $79 - 436\times$ and $137 - 242\times$ faster in LAN and WAN, respectively. Our PPNN protocols even outperform several 3PC protocols, including SecureNN, ABY³, and Falcon [57, 76, 77], boasting speeds up to $52\times$ in LAN and $37\times$ in WAN.

Figure 10 presents our scalability experiments. Our protocol’s training times scale linearly with n and sub-linearly with d . SecureML only provides results for the client-aided LAN setting.

8.5 Secure Inference

In this section, we evaluate the efficacy of our protocol for secure inference using the MNIST dataset [51]. Our HAWK_{Single} protocol is better suited for secure inference than HAWK_{Multi} due to its reduced online cost. The cost of secure inference using HAWK_{Single} for both logistic regression and neural networks is shown in Table 6, alongside a comparison with SecureML [58]. The neural network architecture referenced is the same as discussed in Section 8.4.

SecureML [58] only presents the inference times for their server-aided protocol. For neural networks, they only report inference

times using the *square* activation function. While the *square* activation is faster than ReLU, it compromises slightly on accuracy, as highlighted in SecureML [58]. Our client-aided protocol has a higher online cost for dot product operation because matrix multiplication isn’t feasible [58]. Moreover, we employ the ReLU activation function, ensuring the preservation of plaintext accuracy. In terms of secure inference, our HAWK_{Single} protocol’s online phase is up to $2.5\times$ faster than SecureML for both logistic regression and neural networks. Additionally, the offline phase of HAWK_{Single} outperforms SecureML across most scenarios.

9 CONCLUSION

In this work, we present novel lookup table protocols for PPML that achieve both accuracy and efficiency in activation function computation. We present HAWK_{Single}, offering leakage-free security, and HAWK_{Multi}, enabling table reuse at the cost of leakage which preserves $d_{\mathcal{X}}$ -privacy. Our PPML protocols for logistic regression and neural networks demonstrate significant performance gains, reaching up to $688\times$ faster training for neural networks than SecureML [58] with plaintext accuracy. Beyond activation functions, our approach applies to various MPC applications, paving the way for wider adoption of PPML and secure collaborative data analysis.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] Nitin Agrawal, Ali Shahin Shamsabadi, Matt J. Kusner, and Adrià Gascón. 2019. QUOTIENT: Two-Party Secure Neural Network Training and Prediction (CCS '19). 1231–1247.
- [2] Joshua Allen, Bolin Ding, Janardhan Kulkarni, Harsha Nori, Olga Ohrimenko, and Sergey Yekhanin. 2019. An algorithmic framework for differentially private data analysis on trusted processors. *Advances in Neural Information Processing Systems* 32 (2019).
- [3] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-Indistinguishability: Differential Privacy for Location-Based Systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. Association for Computing Machinery, 901–914.
- [4] Yoshinori Aono, Takuya Hayashi, Le Trieu Phong, and Lihua Wang. 2016. Scalable and Secure Logistic Regression via Homomorphic Encryption. In *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy (CODASPY '16)*. 142–144. <https://doi.org/10.1145/2857705.2857731>
- [5] Gilad Asharov, Ilan Komargodski, Wei-Kai Lin, Kartik Nayak, Enoch Peserico, and Elaine Shi. 2020. OptORAMA: optimal oblivious RAM. In *Advances in Cryptology—EUROCRYPT 2020: 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10–14, 2020, Proceedings, Part II* 30. Springer, 403–432.
- [6] Kartik Audhkhasi, Osonde Osoba, and Bart Kosko. 2013. Noise benefits in back-propagation and deep bidirectional pre-training. *Proceedings of the IJCNN*, 1–8.
- [7] Alessandro N. Baccarini, Marina Blanton, and Chen Yuan. 2023. Multi-Party Replicated Secret Sharing over a Ring with Applications to Privacy-Preserving Machine Learning. *Proc. Priv. Enhancing Technol.* 2023 (2023), 608–626.
- [8] Donald Beaver. 1991. Efficient multiparty protocols using circuit randomization. In *Annual International Cryptology Conference*. Springer, 420–432.
- [9] Barry Becker. 1994. Adult Data Set. <https://archive.ics.uci.edu/ml/datasets/Adult>.
- [10] Amos Beimel, Yuval Ishai, and Tal Malkin. 2004. Reducing the Servers' Computation in Private Information Retrieval: PIR with Preprocessing. *Journal of Cryptology* 17 (2004), 125–151.
- [11] C. M. Bishop. 1995. Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation* 7 (1995), 108–116.
- [12] Peter Bogtoft, Ivan Damgård, Thomas Jakobsen, Kurt Nielsen, Jakob Pagter, and Tomas Toft. 2006. A Practical Implementation of Secure Auctions Based on Multiparty Integer Computation. In *Financial Cryptography and Data Security*. Springer Berlin Heidelberg, 142–147.
- [13] Léon Bottou and Olivier Bousquet. 2008. The Tradeoffs of Large Scale Learning. In *Advances in Neural Information Processing Systems*. 161–168.
- [14] Elette Boyle, Niv Gilboa, and Yuval Ishai. 2015. Function Secret Sharing. In *Advances in Cryptology - EUROCRYPT 2015*, Elisabeth Oswald and Marc Fischlin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 337–367.
- [15] Elette Boyle and Moni Naor. 2016. Is there an oblivious RAM lower bound?. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*. 357–368.
- [16] Paul Bunn and Rafail Ostrovsky. 2007. Secure Two-Party k-Means Clustering. In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07)*. Association for Computing Machinery, 486–497.
- [17] Ran Canetti. 1998. Security and Composition of Multi-party Cryptographic Protocols. *Cryptology ePrint Archive*, Report 1998/018.
- [18] Ran Canetti. 2000. Security and composition of multiparty cryptographic protocols. *Journal of Cryptology* 13 (2000), 143–202.
- [19] R. Canetti. 2001. Universally composable security: a new paradigm for cryptographic protocols. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. 136–145.
- [20] T.-H. Hubert Chan, Kai-Min Chung, Bruce Maggs, and Elaine Shi. 2022. Foundations of Differentially Oblivious Algorithms. *J. ACM* 69, 4 (Aug. 2022), 1–49. <https://doi.org/10.1145/3555984>
- [21] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. 2013. Broadening the Scope of Differential Privacy Using Metrics. In *Privacy Enhancing Technologies*, Emiliano De Cristofaro and Matthew Wright (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 82–102.
- [22] Jack Crawford, Craig Gentry, S. Halevi, Daniel Platt, and V. Shoup. 2018. Doing Real Work with FHE: The Case of Logistic Regression. *Proceedings of the 6th Workshop on Encrypted Computing & Applied Homomorphic Cryptography* (2018).
- [23] Wei Dai et al. 2019. Crypto++. <https://www.cryptopp.com/>.
- [24] Ivan Damgård, Jesper Buus Nielsen, Michael Nielsen, and Samuel Ranellucci. 2017. The TinyTable Protocol for 2-Party Secure Computation, or: Gate-Scrambling Revisited. In *Advances in Cryptology – CRYPTO 2017*.
- [25] Ivan Damgård and Rune Thorbek. 2008. Efficient Conversion of Secret-shared Values Between Different Fields. *IACR Cryptology ePrint Archive* 2008 (01 2008), 221.
- [26] Daniel Demmler, Thomas Schneider, and Michael Zohner. 2015. ABY - A Framework for Efficient Mixed-Protocol Secure Two-Party Computation. In *NDSS*.
- [27] G. Dessouky, F. Koushanfar, A. Sadeghi, T. Schneider, Shaza Zeitouni, and Michael Zohner. 2017. Pushing the Communication Barrier in Secure Computation using Lookup Tables. *IACR Cryptol. ePrint Arch.* (2017).
- [28] Jack Doerner and Abhi Shelat. 2017. Scaling ORAM for Secure Computation. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Dallas Texas USA, 523–535. <https://doi.org/10.1145/3133956.3133967>
- [29] Nathan Dowlin, Ran Gilad-Bachrach, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. 2016. *CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy*. Technical Report.
- [30] Wenliang Du, Yunghsiang Han, and Shigang Chen. 2004. Privacy-Preserving Multivariate Statistical Analysis: Linear Regression And Classification. (04 2004). <https://doi.org/10.1137/1.9781611972740.21>
- [31] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3–4 (Aug. 2014), 211–407. <https://doi.org/10.1561/0400000042>
- [32] Natasha Fernandes. 2022. Differential privacy for metric spaces: information-theoretic models for privacy and utility with new applications to metric domains. (10 2022). <https://doi.org/10.25949/21365637.v1>
- [33] R. A. FISHER. 1936. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics* 7 (1936). <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- [34] A. Gascón, Philipp Schoppmann, B. Balle, Mariana Raykova, Jack Doerner, Samee Zahur, and David Evans. 2016. Secure Linear Regression on Vertically Partitioned Datasets. *IACR Cryptol. ePrint Arch.* (2016).
- [35] O. Goldreich, S. Micali, and A. Wigderson. 1987. How to Play ANY Mental Game (STOC '87).
- [36] Oded Goldreich and Rafail Ostrovsky. 1996. Software protection and simulation on oblivious RAMs. *J. ACM* 43, 3 (May 1996), 431–473. <https://doi.org/10.1145/233551.233553>
- [37] S Dov Gordon, Jonathan Katz, Vladimir Kolesnikov, Fernando Krell, Tal Malkin, Mariana Raykova, and Yevgeniy Vahlis. 2012. Secure two-party computation in sublinear (amortized) time. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 513–524.
- [38] Gaël Guennebaud, Benoît Jacob, et al. 2010. Eigen v3. <http://eigen.tuxfamily.org>.
- [39] Suyog Gupta, Ankur Agrawal, Kailash Gopalakrishnan, and Pritish Narayanan. 2015. Deep Learning with Limited Numerical Precision. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML '15)*. 1737–1746.
- [40] Isabelle Guyon. 2003. Arcene Data Set. <https://archive.ics.uci.edu/ml/datasets/Arcene>.
- [41] Isabelle Guyon. 2003. Gisette Data Set. <https://archive.ics.uci.edu/ml/datasets/Gisette>.
- [42] Isabelle Guyon, Steve Gunn, Asa Ben-Hur, and Gideon Dror. 2005. Result Analysis of the NIPS 2003 Feature Selection Challenge. In *NIPS 17*, L. K. Saul, Y. Weiss, and L. Bottou (Eds.). 545–552.
- [43] Darrel Hankerson and Alfred Menezes. 2005. *Elliptic Curve Discrete Logarithm Problem*. Springer US, Boston, MA, 186–189.
- [44] Carmit Hazay and Muthuramakrishnan Venkatasubramanian. 2017. Scalable Multi-party Private Set-Intersection. In *Public-Key Cryptography – PKC 2017*, Serge Fehr (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 175–203.
- [45] Geetha Jagannathan and Rebecca N. Wright. 2005. Privacy-Preserving Distributed k-Means Clustering over Arbitrarily Partitioned Data. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05)*. Association for Computing Machinery, 593–599.
- [46] Chiraag Juvekar, Vinod Vaikuntanathan, and Anantha Chandrakasan. 2018. GAZELLE: A Low Latency Framework for Secure Neural Network Inference. In *Proceedings of the 27th USENIX Conference on Security Symposium (SEC'18)*. 1651–1668.
- [47] Seny Kamara, Payman Mohassel, and Mariana Raykova. 2011. Outsourcing Multi-Party Computation. *Cryptology ePrint Archive*, Report 2011/272.
- [48] Marcel Keller, Emmanuela Orsini, Dragos Rotaru, Peter Scholl, Eduardo Soria-Vazquez, and Srinivas Vivek. 2017. Faster Secure Multi-Party Computation of AES and DES Using Lookup Tables. *Cryptology ePrint Archive*, Report 2017/378.
- [49] Miran Kim, Yongsoo Song, Shuang Wang, Xia Yuhou, and Xiaoqian Jiang. 2017. Secure Logistic Regression Based on Homomorphic Encryption: Design and Evaluation. *JMIR Medical Informatics* (08 2017).
- [50] Kasper Green Larsen and Jesper Buus Nielsen. 2018. Yes, there is an oblivious RAM lower bound!. In *Annual International Cryptology Conference*. Springer, 523–542.
- [51] Yann LeCun, Corinna Cortes, and CJ Burges. 2010. MNIST handwritten digit database. *ATT Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist> 2 (2010).

- [52] Yehuda Lindell and Benny Pinkas. 2000. Privacy Preserving Data Mining. In *Advances in Cryptology – CRYPTO 2000*. Springer Berlin Heidelberg, 36–54.
- [53] Jian Liu, Mika Juuti, Yao Lu, and N. Asokan. 2017. Oblivious Neural Network Predictions via MiniONN Transformations. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*. 619–631. <https://doi.org/10.1145/3133956.3134056>
- [54] Kerry Maletsky. 2020. RSA vs ECC Comparison for Embedded Systems. <https://ww1.microchip.com/downloads/en/DeviceDoc/00003442A.pdf>.
- [55] Sahar Mazloom and S. Dov Gordon. 2018. Secure Computation with Differentially Private Access Patterns. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Toronto Canada, 490–507. <https://doi.org/10.1145/3243734.3243851>
- [56] Paulius Mikičevičius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. 2018. Mixed Precision Training. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=1gs9JgRZ>
- [57] Payman Mohassel and Peter Rindal. 2018. ABY3: A mixed protocol framework for machine learning. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 35–52.
- [58] Payman Mohassel and Yupeng Zhang. 2017. SecureML: A System for Scalable Privacy-Preserving Machine Learning. 19–38.
- [59] Alan F. Murray and Peter J. Edwards. 1994. Enhanced MLP Performance and Fault Tolerance Resulting from Synaptic Weight Noise During Training. *IEEE Transactions on Neural Networks* 5 (1994), 792–802.
- [60] Muhammad Naveed, Seny Kamara, and Charles V Wright. 2015. Inference attacks on property-preserving encrypted databases. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 644–655.
- [61] Giuseppe Persiano and Kevin Yeo. 2023. Lower bound framework for differentially private and oblivious data structures. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 487–517.
- [62] Benny Pinkas and Tzachy Reinman. 2010. Oblivious RAM revisited. In *Advances in Cryptology—CRYPTO 2010: 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15–19, 2010. Proceedings 30*. Springer, 502–519.
- [63] Deevashwer Rathee, Mayank Rathee, Rahul Kranti Kiran Goli, Divya Gupta, Rahul Sharma, Nishanth Chandran, and Aseem Rastogi. 2021. SiRnn: A Math Library for Secure RNN Inference. *2021 IEEE Symposium on Security and Privacy (SP)* (2021), 1003–1020.
- [64] M. Sadeq Riaz, Christian Weinert, Oleksandr Tkachenko, Ebrahim M. Songhori, Thomas Schneider, and Farinaz Koushanfar. 2018. Chameleon: A Hybrid Secure Computation Framework for Machine Learning Applications. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIACCS '18)*. 707–721. <https://doi.org/10.1145/3196494.3196522>
- [65] Ashish P. Sanil, Alan F. Karr, Xiaodong Lin, and Jerome P. Reiter. 2004. Privacy Preserving Regression Modelling via Distributed Computation. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*. Association for Computing Machinery, 677–682. <https://doi.org/10.1145/1014052.1014139>
- [66] Zhiwei Shang, Simon Oya, Andreas Peter, and Florian Kerschbaum. 2021. Obfuscated access and search patterns in searchable encryption. *arXiv preprint arXiv:2102.09651* (2021).
- [67] Elaine Shi, T H Hubert Chan, Emil Stefanov, and Mingfei Li. 2011. Oblivious RAM with $O((\log N)^3)$ worst-case cost. In *Advances in Cryptology—ASIACRYPT 2011: 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4–8, 2011. Proceedings 17*. Springer, 197–214.
- [68] Reza Shokri and Vitaly Shmatikov. 2015. Privacy-Preserving Deep Learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. 1310–1321. <https://doi.org/10.1145/2810103.2813687>
- [69] Aleksandra B. Slavkovic, Yuval Nardi, and Matthew M. Tibbits. 2007. Secure Logistic Regression of Horizontally and Vertically Partitioned Distributed Databases. In *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops (ICDMW '07)*. IEEE Computer Society, 723–728. <https://doi.org/10.1109/ICDMW.2007.84>
- [70] Emil Stefanov, Marten van Dijk, Elaine Shi, T-H Hubert Chan, Christopher Fletcher, Ling Ren, Xiangyao Yu, and Srinivas Devadas. 2018. Path ORAM: an extremely simple oblivious RAM protocol. *Journal of the ACM (JACM)* 65, 4 (2018), 1–26.
- [71] Xiao Sun, Jungwook Choi, Chia-Yu Chen, Naigang Wang, Swagath Venkataramani, Vijayalakshmi (Viji) Srinivasan, Xiaodong Cui, Wei Zhang, and Kailash Gopalakrishnan. 2019. Hybrid 8-bit Floating Point (HFP8) Training and Inference for Deep Neural Networks. In *Advances in Neural Information Processing Systems 32*. 4900–4909.
- [72] Jaideep Vaidya, Hwanjo Yu, and Xiaoqian Jiang. 2008. Privacy-preserving SVM classification. *Knowledge and Information Systems* 14 (02 2008), 161–178. <https://doi.org/10.1007/s10115-007-0073-7>
- [73] Sameer Wagh. 2022. Pika: Secure Computation using Function Secret Sharing over Rings. *Proceedings on Privacy Enhancing Technologies* 2022, 4 (oct 2022), 351–377. <https://doi.org/10.56553/popets-2022-0113>
- [74] Sameer Wagh. 2022. Pika: Secure Computation using Function Secret Sharing over Rings. *Proceedings on Privacy Enhancing Technologies* 2022, 4 (Oct. 2022), 351–377. <https://doi.org/10.56553/popets-2022-0113>
- [75] Sameer Wagh, Paul Cuff, and Prateek Mittal. 2018. Differentially Private Oblivious RAM. *Proceedings on Privacy Enhancing Technologies* 2018, 4 (Aug. 2018), 64–84. <https://doi.org/10.1515/popets-2018-0032>
- [76] Sameer Wagh, Divya Gupta, and Nishanth Chandran. 2019. SecureNN: Efficient and Private Neural Network Training. In *Privacy Enhancing Technologies Symposium*. (PETS 2019).
- [77] Sameer Wagh, Shruti Tople, Fabrice Benhamouda, Eyal Kushilevitz, Prateek Mittal, and Tal Rabin. 2021. Falcon Honest-Majority Maliciously Secure Framework for Private Deep Learning. *Proceedings on Privacy Enhancing Technologies* 2021, 1 (2021).
- [78] Naigang Wang, Jungwook Choi, Daniel Brand, Chia-Yu Chen, and Kailash Gopalakrishnan. 2018. Training Deep Neural Networks with 8-bit Floating Point Numbers. In *Advances in Neural Information Processing Systems 31*. 7675–7684.
- [79] Xiao Shaun Wang, Yan Huang, TH Hubert Chan, Abhi Shelat, and Elaine Shi. 2014. SCORAM: oblivious RAM for secure computation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 191–202.
- [80] Wenliang Du and M. J. Atallah. 2001. Privacy-preserving cooperative scientific computations. In *Proceedings. 14th IEEE Computer Security Foundations Workshop, 2001*. 273–282.
- [81] Shuang Wu, Tadanori Teruya, and Junpei Kawamoto. 2013. Privacy-preservation for Stochastic Gradient Descent Application to Secure Logistic Regression. In *The 27th Annual Conference of the Japanese Society for Artificial Intelligence*.
- [82] Pieter Wuille, Tim Ruffing, et al. 2013. libsecp256k1. <https://github.com/bitcoin-core/secp256k1>.
- [83] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. *Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms*.
- [84] Hwanjo Yu, Jaideep Vaidya, and Xiaoqian Jiang. 2006. Privacy-Preserving SVM Classification on Vertically Partitioned Data. In *Advances in Knowledge Discovery and Data Mining*. 647–656.
- [85] Samee Zahur, Xiao Wang, Mariana Raykova, Adrià Gascón, Jack Doerner, David Evans, and Jonathan Katz. 2016. Revisiting Square-Root ORAM: Efficient Random Access in Multi-party Computation. 218–234. <https://doi.org/10.1109/SP.2016.21> ISSN: 2375-1207.

A MACHINE LEARNING

Stochastic Gradient Descent (SGD) is a commonly used iterative algorithm to find the local minimum of a function. We use SGD to minimize the cross-entropy cost function and learn \mathbf{w} . The coefficient vector \mathbf{w} is first initialized to 0 or random values. Then in each iteration, we randomly select a data point (\mathbf{x}_i, y_i) and update \mathbf{w} as $\mathbf{w}_j = \mathbf{w}_j - \alpha \frac{\partial C_i(\mathbf{w})}{\partial w_j}$, where α is the learning rate that controls the step towards the minimum taken in each iteration. Substituting the cost function, we get the update rule: $\mathbf{w}_j = \mathbf{w}_j - \alpha(f(\mathbf{x}_i, \mathbf{w}) - y_i)x_{ij}$.

Batching. Instead of updating \mathbf{w} using a single data sample in each iteration, the dataset is divided into random batches of size $|B|$ and \mathbf{w} is updated by averaging the partial derivative of the samples in a batch. With batching, the vectorized form of the update rule for the k th batch is given as: $\mathbf{w} = \mathbf{w} - \alpha \frac{1}{|B|} \mathbf{X}_k^T (f(\mathbf{X}_k \times \mathbf{w}) - \mathbf{Y}_k)$. One pass through the training samples is called an epoch.

Learning Rate. The learning rate α determines the step size of the SGD towards the minimum. A large α corresponds to larger steps, which may result in divergence due to overstepping the minimum. In contrast, a small α may increase the total iterations required.

Termination. When the cost does not decrease significantly for several iterations, SGD is considered to have converged to a minimum, and we can terminate. We can then use the learned coefficient vector \mathbf{w} for inference.

Algorithm 10 Logistic Regression: $\Pi_{\text{Logistic}(P_0, P_1)}$ **Input:** $\langle X \rangle, \langle Y \rangle$ **Output:** \mathbf{w} **Common Randomness:** $\langle U \rangle, \langle V \rangle, \langle Z \rangle, \langle V' \rangle, \langle Z' \rangle,$

1. Initialize lookup table counters $c_1, \dots, c_m = 0$, where m is the number of clients.
2. For $i \in \{0, 1\}$, P_i computes $\langle E \rangle_i = \langle X \rangle_i - \langle U \rangle_i$. The parties reconstruct E using $\text{Rec}(\langle E \rangle_0, \langle E \rangle_1)$.
3. For $1 \leq j \leq t$
 - 3.1. Parties select a mini-batch denoted by B_j .
 - 3.2. For $i \in \{0, 1\}$, P_i computes $\langle F_j \rangle_i = \langle \mathbf{w} \rangle_i - \langle V \rangle_i$. The parties reconstruct F_j using $\text{Rec}(\langle F_j \rangle_0, \langle F_j \rangle_1)$.
 - 3.3. For $i \in \{0, 1\}$, P_i computes $\langle G_{B_j} \rangle_i = \langle X_{B_j} \rangle_i \times F_j + E_{B_j} \times \langle \mathbf{w} \rangle_i + \langle Z_j \rangle_i - i \cdot E_{B_j} \times F_j$
 - 3.4. For $i \in \{0, 1\}$, P_i calls $\Pi_{\text{Sigmoid}(P_0, P_1)}$ with inputs $(\langle G_{B_j} \rangle_i, L_i^{\text{Sigmoid}})$ and sets the output as $\langle Y_{B_j}^* \rangle_i$.
 - 3.5. For $i \in \{0, 1\}$, P_i computes $\langle D_{B_j} \rangle_i = \langle Y_{B_j}^* \rangle_i - \langle Y_{B_j} \rangle_i$
 - 3.6. For $i \in \{0, 1\}$, P_i computes $\langle F'_j \rangle_i = \langle D_{B_j} \rangle_i - \langle V' \rangle_i$. The parties reconstruct F'_j using $\text{Rec}(\langle F'_j \rangle_0, \langle F'_j \rangle_1)$.
 - 3.7. For $i \in \{0, 1\}$, P_i computes $\langle \Delta \rangle_i = \langle X_{B_j}^T \rangle_i \times F'_j + E_{B_j}^T \times \langle D_{B_j} \rangle_i + \langle Z'_j \rangle_i - i \cdot E_{B_j}^T \times F'_j$
 - 3.8. For $i \in \{0, 1\}$, P_i truncates $\langle \Delta \rangle_i$ element-wise to get $\lfloor \langle \Delta \rangle_i \rfloor$
 - 3.9. For $i \in \{0, 1\}$, P_i updates the coefficient vector as $\langle \mathbf{w} \rangle_i := \langle \mathbf{w} \rangle_i - \frac{\alpha}{|B|} \lfloor \langle \Delta \rangle_i \rfloor$
4. The parties reconstruct \mathbf{w} using $\text{Rec}(\langle \mathbf{w} \rangle_0, \langle \mathbf{w} \rangle_1)$ and output \mathbf{w} .

B REDUCED PRECISION TRAINING

ML algorithms are inherently resilient to error, setting them apart from conventional algorithms that demand precise computations and high dynamic range number representations. High precision computation for training ML algorithms is unnecessary [13]. In fact, introducing minor noise during training can enhance neural network generalization, thereby mitigating over-fitting [6, 11, 59].

A recent trend in neural network training is to use approximate computing techniques to reduce memory requirements and improve computing efficiency [56]. Gupta et al. [39] show that training neural networks using 16-bit fixed-point representation with stochastic rounding can yield accuracy similar to the 32-bit floating-point format. Wang et al. [78] employ a combination of 16-bit and 8-bit floating-point numbers for various deep-learning tasks. Sun et al. [71] present an 8-bit hybrid floating-point format and demonstrate its robustness by training neural networks for various tasks while preserving accuracy.

C PROOFS

THEOREM 5. *The geometric mechanism defined in Definition 4 satisfies $d_{\mathcal{X}}$ -privacy.*

PROOF SKETCH. Let $x, x' \in \mathcal{X}$. Let p_x and $p_{x'}$ denote the probability mass functions of G_x and $G_{x'}$ respectively. Then, for some $z \in \mathcal{Y}$, we have,

$$\begin{aligned} \frac{p_x(z)}{p_{x'}(z)} &= \frac{\frac{1-e^{-\epsilon}}{1+e^{-\epsilon}} \cdot e^{-\epsilon \cdot d_{\mathcal{X}}(x,z)}}{\frac{1-e^{-\epsilon}}{1+e^{-\epsilon}} \cdot e^{-\epsilon \cdot d_{\mathcal{X}}(x',z)}} \\ &= \frac{e^{-\epsilon \cdot d_{\mathcal{X}}(x,z)}}{e^{-\epsilon \cdot d_{\mathcal{X}}(x',z)}} \\ &= e^{\epsilon \cdot (d_{\mathcal{X}}(x',z) - d_{\mathcal{X}}(x,z))} \end{aligned}$$

Using the triangle inequality for the metric $d_{\mathcal{X}}(\cdot, \cdot)$ and as $d_{\mathcal{X}}(x, x') = d_{\mathcal{X}}(x', x)$, we get:

$$\frac{p_x(z)}{p_{x'}(z)} \leq e^{\epsilon \cdot d_{\mathcal{X}}(x, x')}$$

which completes the proof. \square

D LOGISTIC REGRESSION PROTOCOL

We present our complete Logistic Regression protocol in Algorithm 10. To implement logistic regression, we use SecureML's linear regression algorithm and add our Π_{Sigmoid} protocol for computing the Sigmoid function.

We can denote the input matrix for the i th layer as a $|B| \times d_i$ matrix X_i and the coefficient matrix as a $d_{i-1} \times d_i$ matrix W_i . In iteration k , during the forward pass, the input to the first layer X_0 is initialized as batch k of the input data, and X_i , for each layer i , is computed as $X_i = f(X_{i-1} \times W_i)$. During the back propagation, given a cost function $C(W)$, such as the cross-entropy function, we first compute $\Delta_i = \frac{\delta C(W)}{\delta Z_i}$ for each layer i , where $Z_i = X_{i-1} \times W_i$. For the output layer l , Δ_l is computed as: $\Delta_l = \frac{\delta C(W)}{\delta X_l} \odot \frac{\delta f(Z_l)}{\delta Z_l}$; here $\frac{\delta f(Z_l)}{\delta Z_l}$ is the derivative of the activation function and \odot represents element-wise multiplication. Given Δ_l for the output layer, Δ_i for previous layers is computed as: $\Delta_i = (\Delta_{i+1} \times W_i^T) \odot \frac{\delta f(Z_i)}{\delta Z_i}$. The coefficients W_i in layer i are then updated as: $W_i = W_i - \alpha \frac{1}{|B|} (X_i \times \Delta_i)$. We use Rec to denote share reconstruction.

E DATASETS

We use the following datasets in our experiments to compare our results with the previous works [58, 76] and evaluate the scalability of our protocols.

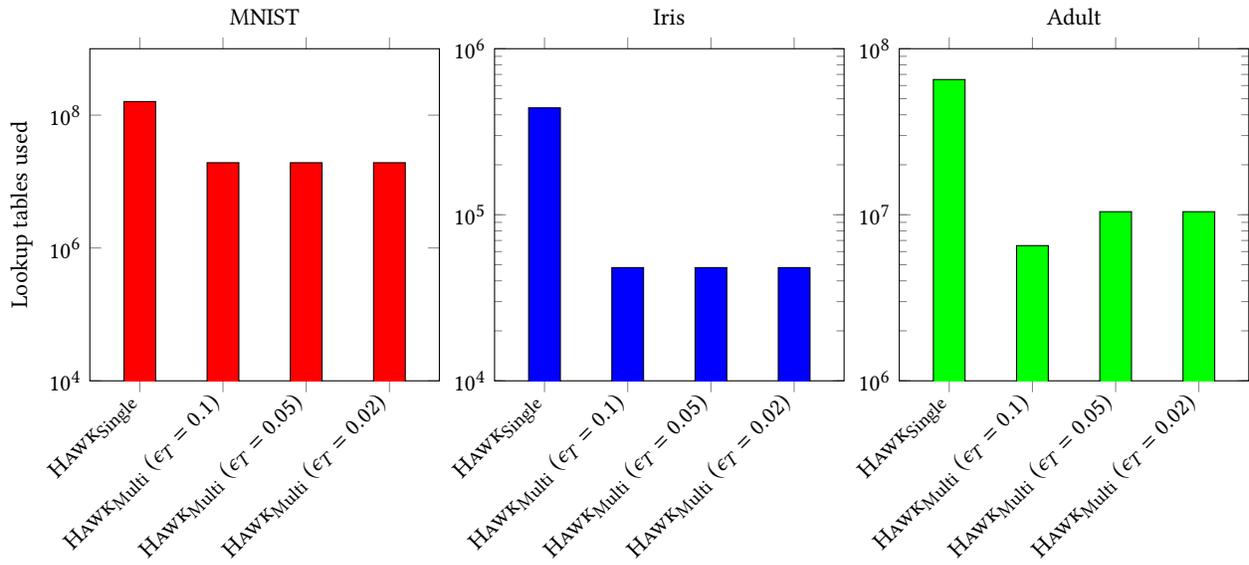


Figure 11: Lookup tables required by our PPNN protocols to achieve a certain accuracy level i.e. 96% for MNIST, 85% for Adult, and 96% for Iris dataset. ($r_{Multi} = 10$).

	Mode	10 ⁰	10 ¹	10 ²	10 ³	10 ⁴	10 ⁵
Pika	LAN	0.0007	0.001	0.006	0.059	0.539	5.329
SIRNN		—	0.115	0.116	0.147	0.289	2.089
HAWKSingle		0.0002	0.0004	0.0004	0.0013	0.0060	0.0482
HAWKMulti		0.0008	0.0016	0.0038	0.0209	0.1855	1.8193
Pika	WAN	0.183	0.195	0.198	0.356	1.375	9.030
SIRNN		—	7.683	7.690	7.816	9.313	32.702
HAWKSingle		0.050	0.059	0.059	0.062	0.066	0.116
HAWKMulti		0.179	0.180	0.185	0.192	0.318	2.160

Table 7: Comparison of HAWK with Pika and SIRNN semi-honest protocols. Time is in seconds and the computation is a sigmoid function on a batch of 10⁰, 10¹, 10², 10³, 10⁴, and 10⁵ in both the LAN and WAN settings.

MNIST. The MNIST dataset [51] contains images of handwritten digits from ‘0’ to ‘9’. The training set consists of 60,000 images, and the test set contains 10,000 images. Each sample consists of 784 features representing a 28 × 28 image with pixel values ranging from 0 to 255.

Gisette. The Gisette dataset [41, 42] is constructed from the MNIST dataset, and the task is to classify the highly confusable digits ‘4’ and ‘9’. The dataset contains 13,500 samples, each with 5,000 features, with values ranging from 0 to 1,000.

Arcene. The Arcene dataset [40, 42] contains mass-spectrometric data, and the task is distinguishing cancer from normal patterns. The dataset contains 900 instances, each with 10,000 features, with values ranging from 0 to 1000.

Fashion MNIST. The dataset contains Zalando’s article images in 10 classes [83]. The training set consists of 60,000 images, and the test set contains 10,000 images. Each sample consists of 784 features with values ranging from 0 to 255. The dataset can be used as a drop-in replacement for MNIST.

Iris. The Iris dataset [33] contains 3 classes of 50 instances, each where each class is a type of iris plant, and each instance consists of 4 features.

Adult. The dataset consists of census data with 48,842 instances, each with 14 attributes. The classification task is to predict if an individual’s income exceeds \$50K/year. The dataset also contains missing values.

F COMPARISON WITH RELATED WORKS

Table 7 provides comparison of our HAWKSingle and HAWKMulti protocols with Pika [74] and SIRNN [63] for sigmoid function computation. Our protocols outperform both Pika and SIRNN in all settings.