



**HAL**  
open science

## **SoK: Verifiable Cross-Silo FL**

Aleksei Korneev, Jan Ramon

► **To cite this version:**

| Aleksei Korneev, Jan Ramon. SoK: Verifiable Cross-Silo FL. 2024. hal-04692331v2

**HAL Id: hal-04692331**

**<https://hal.science/hal-04692331v2>**

Preprint submitted on 10 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# SoK: VERIFIABLE CROSS-SILO FL

---

A PREPRINT

**Aleksei Korneev**  
University of Lille, Inria,  
CNRS, Centrale Lille,  
UMR 9189 - CRISTAL  
aleksei.korneev@inria.fr

**Jan Ramon**  
University of Lille, Inria,  
CNRS, Centrale Lille,  
UMR 9189 - CRISTAL  
jan.ramon@inria.fr

September 9, 2024

## ABSTRACT

Federated Learning (FL) is a widespread approach that allows training machine learning (ML) models with data distributed across multiple devices. In cross-silo FL, which often appears in domains like healthcare or finance, the number of participants is moderate, and each party typically represents a well-known organization. For instance, in medicine data owners are often hospitals or data hubs which are well-established entities. However, malicious parties may still attempt to disturb the training procedure in order to obtain certain benefits, for example, a biased result or a reduction in computational load. While one can easily detect a malicious agent when data used for training is public, the problem becomes much more acute when it is necessary to maintain the privacy of the training dataset. To address this issue, there is recently growing interest in developing verifiable protocols, where one can check that parties do not deviate from the training procedure and perform computations correctly. In this paper, we present a systematization of knowledge on verifiable cross-silo FL. We analyze various protocols, fit them in a taxonomy, and compare their efficiency and threat models. We also analyze Zero-Knowledge Proof (ZKP) schemes and discuss how their overall cost in a FL context can be minimized. Lastly, we identify research gaps and discuss potential directions for future scientific work.

**Keywords** federated learning, FL, cross-silo FL, verification, verifiable protocols, zero knowledge proofs, ZKP

## 1 Introduction

Nowadays, the broad propagation of ML technologies is rapidly increasing. ML applications affect a variety of fields such as medicine, finance, marketing, education, and many others [19, 46, 27, 48]. Many ML approaches rely on a process of training on historical data: a model learns statistical patterns that later allow new predictions to be inferred. However, in some cases, data may contain private or confidential information; therefore, access to such data is limited, and applying ML must be done with extreme caution, either due to an interest in privacy of the data owners (DOs), e.g., individual persons caring about their privacy or companies caring about intellectual property, or for regulatory compliance, e.g., with the General Data Protection Regulations (GDPR).

In FL, multiple DOs, who are also sometimes referred to as clients, can train a model together, possibly under coordination of a central server, by exchanging encrypted messages without revealing their private data. As a result, researchers can benefit from a large amount of shared data and at the same time preserve privacy. However, while preserving privacy in FL allows protecting sensitive information, at the same time it produces an additional challenge in the verification of participants' behavior. Indeed, due to a possibility of malicious actions, it is important to ensure that all calculations are performed correctly even if the used data is private.

FL is often divided into two categories: cross-device and cross-silo. In the cross-device FL setting, data comes from a large number of small and usually anonymous devices with low computational capacities. Anonymity complicates penalizing clients; a single client is free to abort or to violate the procedure at any time. In contrast, in this paper we focus on cross-silo FL where the number of parties is moderate; each party is usually a well-known and large entity

that is expected to cooperate in the entire training process via devices with high computing power. Each party has an incentive to care about its reputation and can be held liable if it is found to be fraudulent. For instance, cross-silo FL appears in the healthcare domain, where DOs are medical centers or hospitals that collaborate to train ML models to improve patient care. Although the need for countermeasures against malicious attacks in such setting could be reduced due to the liability of participants, verification of the calculations is still required to establish confidence in FL’s performance.

Recently, dozens of research works devoted to verifiable FL have been published, proposing methods to ensure the verifiability of the parties’ computations, using different infrastructures and relying on various assumptions. Nonetheless, to the best of our knowledge, verifiability in the context of cross-silo FL has not been thoroughly studied. Chao et al. in [30] studied challenges of cross-silo FL setting in details, but the verifiability property was not taken into account. In [60, 53] authors were focused on verifiability in FL, however features of the cross-silo setting were not considered and protocols’ efficiency was not analysed. Mansouri et al. presented a SoK paper [44] devoted to secure aggregation protocols and included verification in the list of challenges, nevertheless, specific features of cross-silo FL were not in the scope of the paper and an efficiency analysis of verification techniques was not performed. Lastly, in [37, 57] authors studied applications of various ZKP schemes for ML, but these works do not address FL.

In this paper, we provide a systematization of knowledge on verifiable cross-silo FL. Our contributions are summarized as follows:

- to the best of our knowledge, we are the first to conduct a survey on verifiable FL while studying specific challenges of the cross-silo setting;
- we propose a new taxonomy of existing verifiable cross-silo FL protocols while analyzing their efficiency and threat models;
- we perform a comparison of ZKP schemes from the perspective of applying them in cross-silo FL and discuss the influence of parameters such as the ZKP scheme and circuit size on the cost;
- we define future challenges and identify research gaps.

The rest of the paper is organized in the following way: Section 2 introduces some background and notations of FL and cryptographic primitives; Section 3 presents an analysis of existing verifiable cross-silo FL protocols; Section 4 is devoted to a comparison of ZKP schemes from the perspective of their applicability for cross-silo FL, then we describe a storage cost optimization for ZKP-based FL protocols; Section 6 describes challenges and research gaps; and Section 7 concludes the paper.

## 2 Background

### 2.1 FL process

The majority of FL approaches contain two main categories of operations: local operations (both at the side of the clients and the server) and aggregation of clients’ values. Additionally, there are also other operations specific to certain FL protocols, for example, where participants should draw random numbers, exchange cryptographic keys or select a subset of parties to communicate with.

In this paper, we consider both settings where the aggregation is coordinated or performed by a central server and settings where DOs perform the aggregation in a decentralized way.

Some ML algorithms involve running an optimization algorithm, e.g., stochastic gradient descent (SGD). We refer to each iteration of such algorithm as an epoch.

### 2.2 Cross-silo FL properties

While analyzing the suitability of various algorithms for the cross-silo FL setting, we assume that:

1. the number of participants is moderate (at most several thousands);
2. all participants have an incentive to care about their reputation, they may only cheat in a way which can not be detected by others;
3. all participants agree on the model to be trained (type of calculations to be executed);
4. DOs possess computationally sufficiently powerful equipment;

### 2.3 Adversarial attacks on FL

In a survey on FL threats by Rodríguez-Barroso et al. [49] two classes of adversarial attacks were distinguished:

- privacy attacks, whose purpose is to infer sensitive information from the learning process;
- attacks which aim at modifying the behaviour or output of the FL process.

The development of a FL protocol that is resistant to adversarial attacks requires applying a combination of various privacy enhancing technologies (PETs). For example, in order to prevent privacy attacks, authors of state-of-the-art solutions employ differential privacy (DP), multi-party computation (MPC), secure shuffling and Trusted-Execution Environment (TEE) among others.

In this paper, we study verification techniques that allow mitigating the second class of attacks, attacks on the federated learning process, such as data or model poisoning, when an adversary intentionally uses incorrect data or performs computations incorrectly to bias the resulting model. Authors of the considered papers applied commitment schemes, homomorphic hash functions, ZKP schemes and other methods to ensure that the federated model is computed correctly. A detailed analysis of these methods is presented in sections 3 and 4.

### 2.4 Verifiable FL

In the scope of this paper, we rely on a definition of Verifiable FL inspired by [60]:

*Definition (Verifiable FL).* FL is verifiable if selected parties are able to verify that the tasks of all participants are correctly performed without deviation.

Following this definition, in contrast to the survey [53], we only include approaches that at least partly verify computations of the FL process. For example, we do not analyze protocols which are focused only on verification of identity, ownership, or data provenance. We also exclude protocols considered in [44] that aim to prevent model poisoning attacks by analyzing distribution of values submitted by parties. Such methods efficiently mitigate some attacks, but do not allow to verify the correctness of individual computations or of the individual uses of the input data, e.g., an individual outlier input value is infrequent but possibly valid. Moreover, their efficiency depends on the domain and an attacker strength. On the other hand, we do include in our analysis several protocols devoted to verifiable federated private averaging and verifiable cross-device FL since the same verification techniques could be used in the cross-silo FL setting.

### 2.5 Threat models

In the scope of the considered works, authors usually rely on two widely-spread types of threat models: honest-but-curious (a.k.a. semi-honest) and malicious. According to the standard cryptography definitions, an *honest-but-curious* agent does not deviate from the protocol, but keeps a record of the protocol transcript and analyzes it to gain information about other users, while a *malicious* adversary can deviate from the prescribed protocol instructions and follow an arbitrary strategy to obtain greater benefits. However, in the context of FL, authors often adapt these definitions with additional properties. In order to thoroughly analyze miscellaneous flavors of the applied threat models we distinguish the following four categories:

- **honest (or trusted):** always follows the protocol correctly and is trusted with sensitive information;
- **honest-but-curious:** always follows the protocol correctly, but is not trusted with sensitive information;
- **forgery:** may try to forge different data, but otherwise follows the protocol, is not trusted with sensitive information;
- **malicious:** can arbitrary deviate from the protocol and is not trusted with sensitive information.

In the scope of this paper, in order to describe different approaches in a rigorous manner, we specify the robustness of protocols to participant drop-outs, i.e. agents who register to participate but subsequently abandon the protocol, separately from the aforementioned categories of threat models. For example, to report that a protocol is robust against forging and drop-outs, we denote its threat model as "forger + drop".

### 2.6 Blockchain technology

A blockchain is a sequence of blocks, which holds a complete list of transaction records like a conventional public ledger [55]. A transaction could be any action taking place on a blockchain network, for example, a transfer of digital

currency from one party to another. In order to achieve the agreement about the ledger’s state, blockchains rely on consensus mechanisms. For instance, one of the most popular mechanisms is a Proof of work (PoW), where parties, called miners, calculate a hash of the constantly changing block header. When one miner obtains a relevant value, all others must confirm its correctness and add a collection of transactions used for the calculations as a new block.

In certain blockchain settings, parties can deploy on the blockchain code scripts, called smart contracts, that run synchronously on multiple nodes of the blockchain. Smart contracts can be stored in the blockchain and can be automatically executed when certain pre-conditions are met [64].

## 2.7 Commitment scheme

A commitment scheme (CS) is a cryptographic primitive that allows parties to commit to values while keeping them hidden from others [6]. A party cannot modify the value after committing to it, but can later reveal it. CSs have two properties: hiding, meaning that a commitment reveals nothing about the original value, and binding, meaning that a party cannot compute the same commitment from different values, typically due to a computationally hard underlying problem. Lastly, some CSs are also homomorphic, meaning that there are two binary operations  $+$  and  $\cdot$  defined in the domain of original values and the domain of their commitments respectively, such that the following condition holds:  $C(a + b) = C(a) \cdot C(b)$ , where  $a, b$  are values possessed by a party (or parties) and the commitment is represented by  $C$ .

## 2.8 Zero knowledge proof

A Zero-Knowledge Proof (ZKP) is a cryptographic method by which a party called the prover convinces another party, the verifier, about a statement over committed values [21]. A ZKP of a statement should satisfy three properties:

- completeness: if the statement is true, then an honest verifier will be convinced by an honest prover;
- soundness: if the statement is false, an honest verifier will be convinced with at most a negligible probability;
- zero-knowledge: if the statement is true, a verifier is not capable of learning anything but the proven statement itself.

Some ZKP can be turned into non-interactive proofs by replacing the verifier with a random oracle using the Fiat-Shamir heuristic [15]. In the class of non-interactive ZKP, Zero Knowledge Succinct Non-interactive Arguments of Knowledge (zk-SNARKs) are of a particular interest due to the compact proofs relative to the size of the statement and fast verification [45].

In FL, ZKPs can be used to prove that a party correctly performed prescribed computations and, therefore, did not deviate from the FL protocol.

## 3 Verifiable cross-silo FL protocols analysis

In this section, we present a taxonomy of existing verifiable cross-silo FL protocols, analyze the efficiency of verification techniques and threat models, and discuss the impact of the cross-silo setting on verification. In the scope of this section, we refer to the number of clients as  $C$  and to the number of the aggregated vector dimensions as  $D$ .

In order to ensure that a FL protocol is executed correctly, one has to verify both the local computations of the clients, the aggregation, and the local computations of the server(s). We distinguish four categories of different verification techniques and describe each of them in details below. The full taxonomy is presented in Figure 1. Although each approach has specific characteristics, our categories allow observing general design patterns and infer conclusions about their efficiency. For this purpose, we assess computational and communication costs both per client and per server for each method. The comparison of asymptotic complexities of protocols and applied threat models is presented in Table 1 for approaches focused on verifiable aggregation and in Table 2 for approaches focused on verification of local computations. We emphasize that complexity metrics are calculated specifically for the verification overhead and do not reflect computation and communication which is needed even if no verification is performed. Lastly, we assume that public key infrastructure, ML model weights and seeds of PRGs are initialized before the training procedure and do not require a presence of a trusted party.

In order to fairly compare threat models of different approaches, we mapped the threat models described in the considered papers to definitions from Subsection 2.5 and assigned a suitable model ourselves in cases where authors did not explicitly describe it. In such cases, in tables 1 and 2, we use square brackets to denote assigned threat models and collusion markers. Additionally, some articles provided descriptions of multiple threat models, for example, separate

Approach	Computational cost		Communication cost		Threat model		S-C col.	TA	BC
	client	server	client	server	client	server			
VerSA [25]	$O(D)$	$O(CD)$	$O(D)$	$O(CD)$	h-b-c + drop	[forger]	X	X	X
SVeriFL [18]	$O(D)$	$O(CD)$	$O(D)$	$O(CD)$	h-b-c	forger	X	✓	X
Zhang et al. [59]	$O(D)$	$O(C)$	$O(1)$	$O(C)$	[hon]	forger	[X]	X	X
DEVA [54]	$O(CD)$	$O(CD)$	$O(CD)$	$O(CD)$	h-b-c + drop	forger + drop	X	X	X
NIVA [10]	$O(CD)$	$O(CD)$	$O(CD)$	$O(CD)$	[h-b-c + drop†]	[forger + drop]	[X]	X	X
SVFL [41]	$O(D)$	$O(C)$	$O(1)$	$O(C)$	h-b-c	[forger]	X	X	X
Madi et al. [42]	$O(D)$	$O(C)$	$O(1)$	$O(D)$	hon	[forger]	X	X	X
VerifyNet [58]	$O(D)$	$O(CD)$	$O(D)$	$O(CD)$	h-b-c + drop	forger	X	X	X
BytoChain [39]	$O(C + D)$	$O(D)$	$O(1)$	$O(1)$	[mal + drop†]	[forger* + drop]	[X]	X	✓
Fang et al. [14]	$O(C + D)$	$O(1)$	$O(1)$	$O(1)$	[hon + drop†]	[forger* + drop]	[X]	X	✓
VeriFL [24]	$O(C + \frac{D}{E})$	$O(1)$	$O(C)$	$O(1)$	h-b-c + drop	forger	[X]	CRS	X
FedTrust [28]	$O(CD)$	$O(1)$	$O(CD)$	$O(1)$	hon	forger + drop	[X]	X	X
zkDFL [1]	$O(D)$	$O(CD)$	$O(C)$	$O(C)$	[hon + drop†]	[forger*+drop]	[X]	CRS	✓
GOPA [51]	$O(D \log C)$	N/A	$O(D \log C)$	N/A	forger* + drop	N/A	N/A	X	X
zkFL [56]	$O(CD)$	$O(CD)$	$O(1)$	$O(C \log(CD))$	[hon + drop†]	[forger+drop]	[X]	X	X/✓
VFL [16]	$O(D)$	$O(1)$	$O(1)$	$O(1)$	h-b-c	[forger]	[X]	X	X
PVD-FL [61]	$O(D)$	N/A	$O(D)$	N/A	[forger*]	N/A	N/A	X	X

Table 1: A comparison of asymptotic complexities of the aggregation verification overhead per epoch and threat models of verifiable FL protocols. Sections correspond to taxonomy categories (Figure 1). Notations:  $E$  – number of epochs,  $C$  – number of clients,  $D$  – number of vector dimensions, BC – requires a blockchain infrastructure (the (significant) cost of the blockchain is not included in the table), S-C col. – Server-Client collusion to bypass the verification, [ . . . ] – not explicitly described in the paper.

models from privacy and verification perspectives; our table reflects only those related to verification. We note that assigning a threat model for approaches based on blockchain infrastructure poses additional challenges. While clients’ computations in such settings are performed by the clients themselves, aggregation typically occurs as a smart contract. As a result, all miners redundantly compute the aggregation result and a model for the approaches that describe how to obtain it from the aggregation result. In such cases, in Tables 1 and 2, the server’s threat model corresponds to the threat model of miners. However, it is important to mention that blockchain infrastructure itself is also vulnerable to specific threats, which are often not detailed in articles devoted to blockchain-based FL. As an example, Fang et al. [42] explicitly mention that the protocol is robust when 70% of the stake in the system is honest, but for others such assumptions are not always easily visible. In tables 1 and 2, the symbol "\*" corresponds to a threat model applied to a fraction of agents. To sum up, threat models described in our tables for blockchain-based approaches are meaningful only when the assumptions of blockchain infrastructure itself are satisfied.

In tables 1, 2 we also mark whether methods are robust against client drop-out and/or server drop-out. For non-interactive methods (marked with "drop†"), clients only interact once to submit their contribution, which results in a rather trivial robustness against client drop-out: either a client submits the contribution and a later drop-out of the client is not relevant, or the client doesn’t submit the contribution and then doesn’t participate at all. For such approaches, once a client participates, his contribution will be taken into account in the computed aggregate if the protocol finishes. For interactive protocols, in all cases clients need to contribute to some form of "decryption" of the aggregate. Some approaches [51] try to remove the input of the client(s) who dropped out so the set of clients who decrypt is exactly the set of clients who contributed input. In other approaches [58, 25, 54], robustness against client drop-out is achieved by a threshold secret sharing where only part of the clients are needed for this decryption. This decreases security but avoids the need for computations to be rolled back. When we mark the server threat model as robust against drop out, in all cases there are multiple servers (in the case of blockchain based methods multiple miners) of which a certain fraction can drop out without preventing the computation of the final result, e.g., due to a threshold secret sharing scheme among servers or due to the consensus properties of blockchains.

### 3.1 Taxonomy description

**Redundant aggregation (RA) based verification.** This category consists of approaches that require the server to aggregate some redundant values in order to prove that the aggregation of clients’ values is performed correctly. This feature leads to a computational cost of the server to be at least  $O(C)$ . Moreover, some protocols [54, 10, 25] are designed under the assumption that each party has one secret value, therefore a naive scaling of the approach to a FL setting where parties share multi dimensional data would lead to an additional factor  $D$  in the complexity.

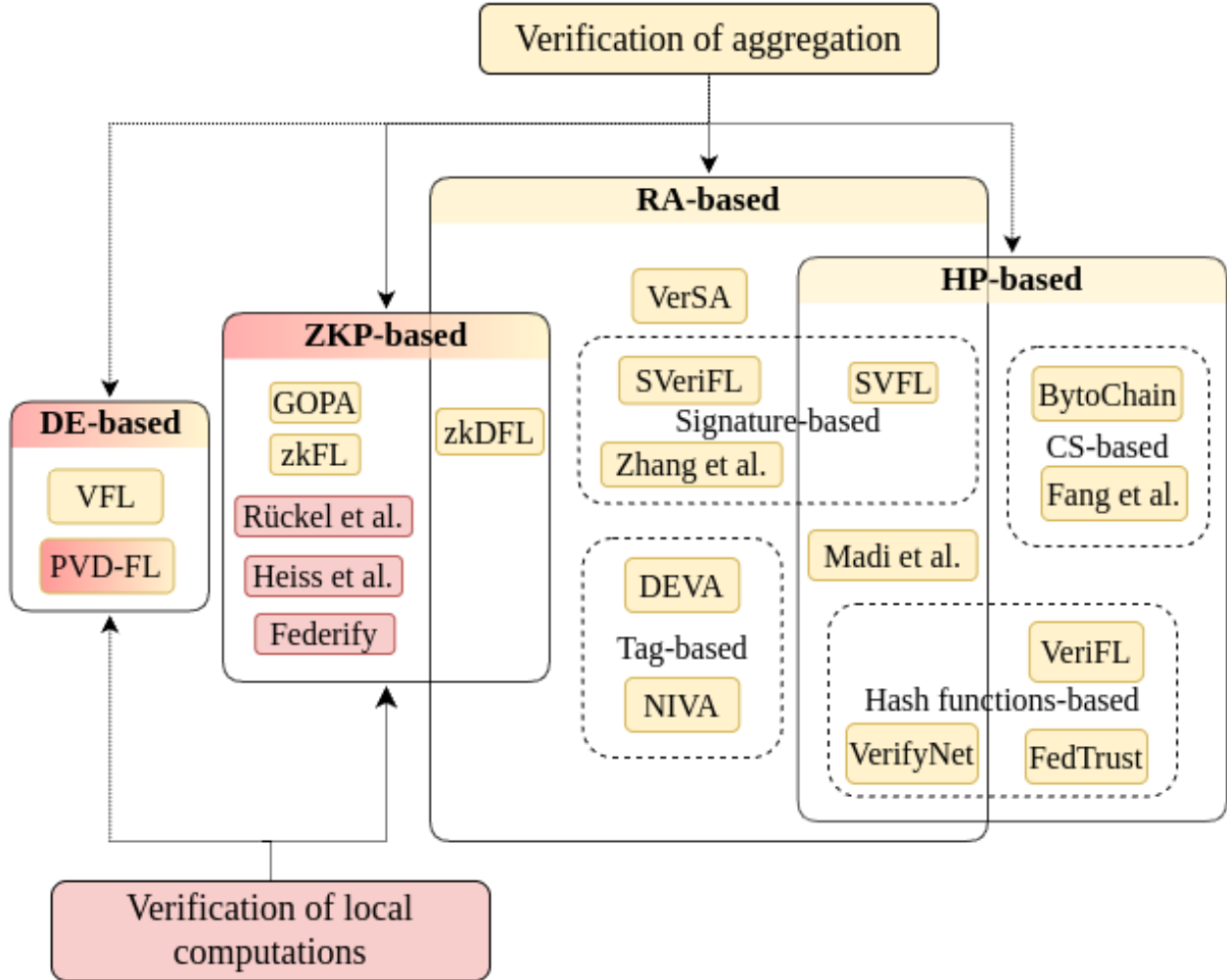


Figure 1: A taxonomy of verifiable cross-silo FL protocols. The red color corresponds to approaches focused on the verification of clients’ computations, the yellow color is used for approaches focused on the aggregation verification.

In [59, 18] authors proposed to check that the result of aggregation is correct by means of cryptographic signature schemes based on bilinear pairings. In both works, the server has to compute a redundant aggregation of signatures. Indeed, such schemes allow ensuring that the aggregation result is obtained from data signed by all other clients, however a malicious server may aggregate arbitrary signed values (e.g., clients’ values from previous epochs) and successfully pass the verification with a fabricated aggregated value. As a result, by relaxing the threat model, [59] achieves a better complexity than other protocols of the same category.

In DEVA [54] and NIVA [10] decentralized FL protocols are proposed, all clients compute tags corresponding to the secret clients’ values which are later used to verify that the aggregation is correct. Both protocols have an additional  $D$  complexity factor due to the assumption that each client has only one secret value. In contrast to many others, DEVA and NIVA present solutions for a setting with multiple servers. Another RA-based approach is VerSA [25]. In VerSA all clients use the same secret vectors to compute a new input from their secret values and later check the consistency of the aggregated result with the aggregation of new inputs. Interestingly, authors provide a comparison with VerifyNet [58] approach (which lies in the intersection of RA- and HP-based categories) showing that despite the same verification asymptotic complexity, the VerSA’s cost is orders of magnitude smaller.

Finally, from the threat model perspective, all RA-based approaches cope with a forger server while considering clients to be honest or honest-but-curious. Additionally, in [18], integrity of data shared by clients is verified, this threat is processed by addition of a trusted authority (TA).

**Homomorphic property (HP) based verification.** This category covers verification techniques which rely on the HP of underlying primitives: hash functions [24, 28] and commitment schemes [39, 14]. The general idea of such protocols

Approach	Computational cost		Communication cost		Threat model		S-C col.	TA	BC
	client	server	client	server	client	server			
Rückel et al. [50]	$O(D)$	$O(1)$	$O(1)$	$O(1)$	[forger + drop <sup>†</sup> ]	[forger* + drop]	✗	CRS	✓
Heiss et al. [26]	$O(D)$	$O(1)$	$O(1)$	$O(1)$	[forger + drop <sup>†</sup> ]	[forger* + drop]	✗	CRS	✓
Federify [32]	$O(D)$	$O(1)$	$O(1)$	$O(1)$	[forger + drop <sup>†</sup> ]	[forger* + drop]	✗	CRS	✓
PVD-FL [61]	$O(D)$	N/A	$O(D)$	N/A	[forger*]	N/A	N/A	✗	✗

Table 2: A comparison of asymptotic complexities of the local computations verification overhead per epoch and threat models of verifiable FL protocols. Sections correspond to taxonomy categories (Figure 1). Notations:  $C$  – number of clients,  $D$  – number of vector dimensions, BC – requires a blockchain infrastructure (the (significant) cost of the blockchain is not included in the table), S-C col. – Server-Client collusion to bypass the verification, [...] – not explicitly described in the paper.

is the following: clients compute hashes/commitments from their data and share them with each other, then all clients may verify the result of aggregation by checking that this result corresponds to the aggregation of hashes/commitments through homomorphism. As a consequence, both computational and communication costs of the server are  $O(1)$  if the ciphertext length does not depend on  $C$  or  $D$ . In order to verify the aggregation of clients’ values, clients have to compute a hash/commitment in  $O(D)$  from their data and to aggregate hashes/commitments from other clients in  $O(C)$ . In blockchain-based protocols clients upload hashes/commitments with a constant communication cost (to be multiplied with costs related to the blockchain infrastructure, see Section 3.2), while in other protocols, where clients have to exchange messages with each other, the communication cost per client is at least  $O(C)$ . Exceptionally, in FedTrust [28] client costs have additional complexity factor  $D$ , as far as the hash is calculated for each component separately.

In VeriFL [24], apart from the main protocol, authors also demonstrated a way to optimize the verification. Leveraging the homomorphic property of the hash scheme and the repetitive nature of FL calculations, authors proposed an amortized verification which allows decreasing the number of hash function calls. Instead of checking that the product of hashes obtained from all clients is equal to the hash of the aggregated value obtained from the server after each epoch, one can draw a set of random coefficients to compute a linear combination of hashes obtained from all clients across multiple epochs and compare it with a hash of a linear combination of aggregated values obtained from the server using the same coefficients. As a result, in Table 1, the client computational cost of VeriFL differs from competing protocols by additional  $\frac{1}{E}$  factor. Notable, since the core idea of all approaches of HP-based category is very similar, we observe that this optimization could be also applied to any of them.

There are also two protocols that lie at the intersection of the RA- and HP-based categories: VerifyNet [58] and SVFL [41]. In VerifyNet, clients share five additional values along with their gradients, which are later aggregated by the server. Aggregated values are shared back with clients so that they can verify the correctness of the gradients aggregation relying on the properties of the homomorphic hash function. The second approach, SVFL, is based on a homomorphic signature scheme. During the training procedure, the server performs a redundant aggregation of signatures and each client runs the verification algorithm based on the HP. Authors considered a malicious threat model for the server, however one should be careful while using signature-based verification techniques due to the concerns described above for similar approaches of the RA-based verification category [18, 59].

**ZKP based verification.** The third category contains approaches which are based on ZKPs. The core principle could be described as follows: a party performs calculations and at the same time computes the proof, which is shared along with the result of calculations; other parties can later run the proof verification algorithm to ensure that the result was computed correctly. In contrast to previous categories, advanced ZKPs allow proving arbitrary computations, therefore such methods are suitable for proving the correctness of both aggregation of clients’ values [56, 1, 51] and local computations [32, 26, 50]. Moreover, recent ZKP schemes provide a proof size that is sublinear in the amount of computations to prove.

In protocols focused on aggregation, authors build on different infrastructures and ZKP schemes. In GOPA [51], authors introduced a decentralized gossip approach where nodes publish proofs of their computations using  $\Sigma$ -protocols. In zkFL [56] authors apply more modern ZKP scheme, Halo2, and provide two versions of the protocol – with a centralized FL setting and a blockchain based one. In zkDFL [1] authors rely on the blockchain infrastructure and the Groth16 scheme. We also observed that authors of zkDFL and zkFL use different techniques to prove that each client value was indeed sent by one of the clients. In zkFL, all clients sign commitments to their local models and the server includes the signature verification in the proof of own computations. Interestingly, with this example one can notice that ZKPs help to eliminate the need for clients to communicate with each other to verify the integrity of aggregated data. In zkDFL, authors proposed to deploy a smart contract which checks that the result of redundant aggregation of local weights hashes performed by server is equal to the sum of hashes uploaded by clients (notably, computational burden of the



aggregation verification is also transferred to the smart contract). Consequently, differences in the settings and chosen ZKP schemes result in different complexity metrics.

In contrast to protocols with a verifiable aggregation, approaches focused on local computations verification [50, 26, 32] have almost identical design: all protocols use the Groth16 scheme to achieve verifiability of clients' local computations in a blockchain based setting. However, authors focus on verification of different ML models: a linear regression model in [50], a naive Bayes classifier in [32] and a feedforward neural network in [26]. In all three approaches, aggregation is performed by a smart contract, therefore the correctness of the aggregation relies on the blockchain security assumptions. One can see that proving, verification and proof size complexities would be the same for the same type of computations, since all three protocols use the same ZKP scheme. Similarly, in all three protocols participants need to have an access to public parameters (including a CRS) which would also require the same amount of memory for their storage. We note that Table 2 does not reflect the complexities of miners which perform computations represented as smart contracts of the blockchain-based approaches.

Additionally, in [50] clients commit to the Merkle root of their private dataset while registering in the system which allows ensuring data integrity within the whole training procedure. To achieve this, clients extend each proof showing that computations are performed with committed data. In [32], ZKPs are also applied to prove properties of local models, i.e., the distance metrics indicating how close the submitted model is to the global one.

Lastly, it is important to mention that there are many research works devoted to proving the correctness of ML calculations without a FL setting [31, 62, 29]. Often, the same verification techniques and developed optimizations could be applied in the context of FL to prove local computations of participants. We believe that the vast majority of such techniques would fit the ZKP-based verification category of our taxonomy.

**Data Embedding based verification.** This category covers protocols, where participants embed additional values into their data before sharing it with untrusted parties; later, the result of calculations performed by an untrusted source is assumed to be correct if the corresponding additional values are computed correctly. The embedding principle leads to an increase in the size of the transmitted data and the complexity of the outsourced calculations, which depends on the size of embedded values, and require more expensive data preprocessing.

One of the best approaches focused on the verifiable aggregation from the complexity perspective is VFL [16]. In VFL clients encode their secret values as a polynomial function and embed an additional point  $(a_i, A)$  before interpolation, where  $a_i$  is a parameter of client  $i$  and  $A$  is obtained from a pseudo random generator (PRG) using  $a_i$  as an input. In order to verify the result of aggregation performed by a malicious server, each client check that the evaluation of the aggregated polynomial function at the point  $a_i$  corresponds to the output of PRG with  $\sum_{i=1}^C a_i$  as an input. In this protocol, transmitted data overhead is negligible, but each client has to compute a costly interpolation and stores large public parameters.

In PVD-FL [61], authors proposed a decentralized verifiable protocol which is based on the verifiable matrix multiplication algorithm. Parties embed random vectors into their data and then check that these vectors were correctly multiplied, as a result, the developed algorithm allows to verify basic operations of ML. In contrast to other approaches surveyed in this paper, in PVD-FL authors aim to show correctness of both aggregation and local models calculation. However, they also mention that their protocol is still vulnerable against poisoning attacks.

### 3.2 Discussion

In this subsection, we discuss how various features of the cross-silo setting impact the development of a verifiable FL protocol. We highlight efficient schemes which cope with cross-silo FL challenges, describe the influence of the threat model choice on the efficiency and security of the protocol, and discuss advantages and disadvantages of the blockchain-based approaches.

Firstly, we describe the link between cross-silo setting characteristics and the efficiency of protocols. One can notice that in tables 1 and 2 there are mainly two parameters determining communication cost:  $C$  and  $D$ . However, there is a large difference in their impact on the complexity. Since the number of participants in cross-silo settings is moderate while ML models typically have large sizes, a dependence on  $D$  is less desirable. Nevertheless, taking into account that clients anyway must send their local models to a server with  $O(D)$  communication cost, the overall FL complexity would become asymptotically worse only in cases when the verification overhead is larger than  $D$ . For example, such as in [54, 10, 28], where the communication cost is  $O(CD)$ .

Secondly, we observe that several approaches rely on a blockchain infrastructure [14, 39, 1, 50, 26]. This strategy offers several advantages. For instance, smart contracts enforce a transparent and verifiable distribution of incentives [50]. The use of smart contracts to perform aggregation also makes the presence of a distinct server unnecessary, thereby replacing a single party trust with blockchain trust guarantees. All blockchain-based approaches are also robust

against limited drop-outs of aggregators, i.e. miners. Table 1 demonstrates that the verification overhead per client is generally smaller for blockchain-based approaches compared to non-blockchain protocols with similar verification techniques; however, there is a significant infrastructure overhead, e.g., costs of the miners, which is a critical limitation within the context of cross-silo FL. The miners have to execute identical calculations, resulting in a tremendous total computational burden across all participants. We note that the miners’ overhead can vary across different blockchains depending on the consensus mechanism: while in Proof-of-Work (PoW) systems like Bitcoin, all miners redundantly execute all computations, in Proof-of-Stake (PoS) systems like Ethereum, transactions are verified by a set of parties called "validators". Although the general idea remains the same, the set of validators is smaller, reducing the overall computational cost. Finally, in cross-silo FL there is typically at least one party interested in obtaining the results of training, thus there is no a strong need for a decentralized infrastructure.

Thirdly, during our analysis of verifiable aggregation protocols, we observed that server has several ways to fabricate the aggregation result, and sometimes a proposed protocol only addresses a portion of them. Specifically, there are three ways in which a forger can manipulate the aggregation of clients’ values: omitting one of the values, replacing one update with arbitrary data or inserting an additional update. However, in RA-based approaches that rely on cryptographic signature schemes, a verifier is only capable of checking that server has not omitted values from other clients and has not inserted additional values in the sum. Nevertheless, a malicious server still may aggregate arbitrary signed values and successfully pass the verification. Therefore, such verification techniques should be employed with caution.

Almost all verifiable aggregation approaches from our analysis rely on honest or honest-but-curious clients’ threat model. The only exceptions with support of malicious clients are: the blockchain based approach BytoChain [39], where a committee of verifiers is tasked to check properties of clients’ uploads, therefore partially mitigating poisoning attacks and free-riding attacks. Indeed, weaker threat models are common in cross-silo FL. However, in real world conditions the use of such threat models is not advisable: a protocol that is resistant to malicious attacks incentivizes participants to avoid cheating, as opposed to a protocol designed with a weaker threat model.

Lastly, our observations indicate that the majority of verifiable aggregation protocols rely on either redundant aggregation mechanisms or leverage the homomorphic properties of cryptographic primitives. However, these techniques are not suitable for proving complex computations, such as those encountered in ML. To address this limitation, researchers employ ZKP schemes, known for the ability to prove arbitrary computations. We examined several ZKP-based protocols developed both to verify aggregation and more complex computation of clients. Furthermore, ZKP schemes offer the flexibility to reinforce the primary proof with additional information, showing the correctness of computed noise, data provenance, or distance metrics. The complexity metrics of ZKP-based FL protocols directly depend on the chosen scheme, however we observe the lack of justifications behind the scheme selection process in the literature. Considering the benefits of ZKPs and their advantage in comparison with other techniques, in the next section we examine existing ZKP schemes with a focus on their applicability in the cross-silo setting.

## 4 ZKP for cross-silo FL

The development of new ZKP schemes has been an active area of research over the past decade. While the surge of new protocols has led to a broad variety of schemes to choose from, it also resulted in additional desirable characteristics, making it challenging to determine the most suitable choice for a specific application. In this context, we discuss the applicability of ZKPs in cross-silo FL and study how to prove calculations in this setting minimizing the cost.

### 4.1 Applicability

In this subsection, we discuss diverse characteristics of ZKP schemes, examining their implications within the context of the cross-silo FL setting. Specifically, we consider the time complexities for proving, verifying and preprocessing, the proof size, the common reference string (CRS) and the commit-and-prove property which some ZKP schemes feature, and the partition into transparent schemes versus trusted setup based schemes.

**Computational complexity of proving and verifying.** In cross-silo FL, the verification of proofs generated during the training procedure presents a significant challenge: each party who wants to ensure the correctness of the protocol needs to execute a verification algorithm in order to check proofs coming from many participants. If the verification algorithm has a linear cost with respect to the size of the computations, the cost for the verifier is proportional to the combined computations of all other parties in the system, which is often too expensive. Furthermore, in scenarios where the protocol is publicly verifiable, e.g. when all proofs are stored on a bulletin board as demonstrated in [51, 10], verification of all proofs performed by an external party after the training procedure would become excessively time-consuming. Hence, we consider it is desirable that the verification time complexity is at most logarithmic in the total amount

of computation to ensure feasibility within the FL setting. In contrast to the verification complexity, the proving complexity is of a slightly less priority since each party has to prove only its own computations once. Consequently, the overhead incurred by executing an algorithm to construct a ZKP is often feasible even if this algorithm is somewhat more expensive.

**Preprocessing computational complexity.** Achieving fast verification and small proofs often comes with a costly preprocessing phase, when the CRS is generated. Since this phase is completed only once before the training process, while the preprocessing complexity is a factor to consider, its impact is generally less critical than for algorithms that are executed within the training loop.

**Proof size.** In a verifiable FL protocol, all parties compute proofs of their local calculations. If the proof size scales linearly with the size of computations, parties need to share data proportional to that computation. Such proofs would require gigabytes to petabytes of storage. In the context of publicly verifiable protocols, this problem becomes even more acute. Therefore, similarly to the requirement for verification cost, it is essential for the proof size to be at most logarithmic.

**Common Reference Strings (CRS).** Numerous SNARKs rely on generating a CRS during the preprocessing phase. There exist two types of CRS: universal, capable of supporting all circuits of bounded size, and custom, applicable only for specific circuits. Given that in FL all parties agree on the model to train, i.e. computations to prove, there is no need to use a scheme with a universal reference string.

In the context of FL, all participants have to access the CRS to compute and verify proofs. The cost of delivering the CRS is determined both by the size of this CRS, and the challenge to broadcast this CRS in a way such that all recipients can trust it. One should take into account that the complexity of the CRS’s size differs depending on the ZKP scheme. Therefore, if the CRS scales linearly to the size of computations to prove and some heavy computations have to be proven, the storage cost will likely become prohibitively large.

**Transparent and trusted setup based setups.** A huge fraction of ZKP schemes depend on trusted setups to generate public parameters. Modern protocols often require only one honest party for this procedure. Since many cross-silo FL protocols are developed under the assumption that at least one honest party is present in the setting, applying such schemes would fit the threat model of these protocols. At the same time, there exist transparent ZKP schemes that eliminate the need for a trusted setup. As a result, integrating such schemes into FL protocol would allow reliance on a more robust threat model.

**Commit-and-prove.** Some ZKP schemes maintain the commit-and-prove property, meaning that the prover can commit to certain data and later prove computations involving this data. As a result, a verifier will be able to check that data has not changed across multiple proofs. In the context of FL, this property is useful to demonstrate that DOs have not changed data during the training. Such a procedure mitigates the risk of one-shot data poisoning attacks and narrows capabilities of potential attackers.

Following the discussion above, in Table 3, we present an asymptotic comparison of various ZKP schemes with at most logarithmic size proofs relative to the size of computations. Our comparison shows that while numerous schemes offer small proof sizes and fast verification, only DARK-based approaches (e.g., SuperSonic, Dew) maintain a constant size of public parameters at the same time.

## 4.2 Storage cost optimization

In the previous subsection, we highlighted that one of the more important criteria for designing a verifiable FL protocol is the total communication cost of ZKP schemes and the associated cost to store the proofs. If parties perform more computations than they want to include in a single proof, they can distribute their computations over multiple proofs. In this subsection we first study what is the best granularity of the proofs under different ZKP schemes and then discuss an alternative idea to prove computations based on recursive proof composition.

As in a ML algorithm the same operations are often repeated many times, e.g., for different data or for different epochs in iterative algorithms, we assume that one can produce a ZKP for all computations by repeatedly proving correct evaluation of a single circuit. Then, we introduce a parameter  $k$  representing how many evaluations of this circuit are included per ZKP. In order to understand the effect of grouping less or more computations together in a single proof, we define a function to compute the storage cost as a function of  $k$  and then find the optimal number  $k$  of grouped circuits to prove, i.e., we find the  $k$  with minimum function value.

To begin with, we introduce the following notations:

- $c \in \mathbb{N}$  – the size of the minimal circuit in bits, i.e., the size of the smallest circuit so that the complete algorithm can be represented as a repetition of that circuit;

Scheme	Parameters size	Proving	Verification	Proof Size
Dory [38]	$O( C )$	$O( C )$	$O(\log C )$	$O(\log C )$
Gemini (space-efficient) [8]	$O( C )$	$O( C \log^2 C )$	$O(\log C )$	$O(\log C )$
Gemini (time-efficient) [8]	$O( C )$	$O( C )$	$O(\log C )$	$O(\log C )$
SuperSonic [12], DARK-fix [2]	$O(1)$	$O( C \log C )$	$O(\log C )$	$O(\log C )$
BCCGP [7]	$O( C )$	$O( C )$	$O( C )$	$O(\log C )$
Bulletproofs [11]	$O( C )$	$O( C )$	$O( C )$	$O(\log C )$
Compressed $\Sigma$ -protocol [3]	$O( C )$	$O( C )$	$O( N )$	$O(\log( C ))$
Groth16 [23]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
Sonic [43]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
GGPR [20]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
Pinocchio [47]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
PLONK [17]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
vnTinyRAM [5]	$O( C \log C )$	$O( C \log^2 C )$	$O( N )$	$O(1)$
Mirage [33]	$O( C )$	$O( C \log C )$	$O( N )$	$O(1)$
Behemoth [52]	$O( C )$	$O( C ^3\log C )$	$O( N )$	$O(1)$
Dew [2]	$O(1)$	$O( C ^2)$	$O(\log C )$	$O(1)$

Table 3: Asymptotic comparison of ZKP schemes with logarithmic and constant proof size complexity.  $C$  is the computation expressed as a circuit,  $|C|$  is the number of gates in the circuit,  $|N|$  is the length of inputs.

- $n \in \mathbb{N}$  – the number of circuits to prove;
- $1 \leq k \in \mathbb{N} \leq n$  – the number of circuits grouped for one proof;
- $\delta : \mathbb{N}^3 \rightarrow \mathbb{N}$  – the function that, for a certain ZKP scheme, gives the total communication cost  $\delta(k; c, n)$  in bits to prove an FL algorithm consisting of  $n$  evaluations of a circuit of size  $c$ , where proofs are given for groups of  $k$  circuits;
- $\lambda : \mathbb{N} \rightarrow \mathbb{N}$  – the function that returns the proof size  $\lambda(c)$  required to prove a circuit of size  $c$  in bits for a certain ZKP scheme;
- $\psi : \mathbb{N} \rightarrow \mathbb{N}$  – the function that gives the size of the public parameters  $\psi(c)$  required to prove circuits of size  $c$  in bits for a certain ZKP scheme.

We also note that the the cost per bit for the distribution and storage of proofs may be different than the cost per bit for the generation and distribution of public parameters, hence we scale  $\psi$  and  $\lambda$  with constants  $a$  and  $b$  to reflect the appropriate weights of a certain ZKP scheme.

Then, for an arbitrary ZKP scheme we can define  $\delta$  in the following way:

$$\delta(k; c, n) = a\psi(kc) + \frac{nb\lambda(kc)}{k} \quad (1)$$

One can see that if  $k = 1$ , then the storage cost is as in a classic verifiable FL protocol without the storage optimization.

One can notice that in Table 3 there are ZKP schemes with the same asymptotic complexities of proof and parameters size. Since in the scope of this optimization, such schemes have the same optimal value of  $k$ , we group schemes with the same complexities together and describe each group below.

#### Group 1: Constant parameters size and constant proof size.

We define the function  $\delta_1$  by instantiating  $\psi$  and  $\lambda$  with corresponding asymptotic complexities of the Group 1 in the definition of  $\delta$ :

$$\delta_1(k; c, n) = a + \frac{nb}{k}$$

One can see that in the simple case when a ZKP scheme (for example, the Dew scheme [2]) has constant parameters and proof size, the function tends to a minimum with the increase of  $k$ , thus the optimal  $k = n$ .

**Group 2: Constant parameters size and logarithmic proof size.**

Several approaches of this group can be found in Table 3: SuperSonic [12] and its improved version called DARK-fix [2]. In order to find an optimal  $k$  for this group, we define  $\delta_2$  similarly to the  $\delta_1$  using the logarithmic proof size complexity:

$$\delta_2(k; c, n) = a + \frac{nb \cdot \log(ck)}{k}$$

Since for  $k, c > 1$  the function is decreasing, the minimum is reached at  $k = n$ .

**Group 3: Linear parameters size and constant proof size.** In table 3, there are many approaches with characteristics of the Group 3: Groth16 [23], Sonic [43], GGPR [20], Pinocchio [47], PLONK [17], Mirage [33], Behemoth [52]. We define  $\delta_3$  as follows:

$$\delta_3(k; c, n) = ack + \frac{nb}{k}$$

We can approximately find the minimum of this function on integers by taking the derivative of the corresponding function on the reals:

$$\frac{d\delta_3(k; c, n)}{dk} = ac - \frac{nb}{k^2}$$

The minimum value of  $\delta_3$  is  $k_e = \sqrt{\frac{nb}{ac}}$ . Since  $k_e$  could be less than one, the optimal  $k$  is  $\min(\max(k_e, 1), n)$ . Note, that in practice one has to check which of the two values  $\lfloor k_e \rfloor$  or  $\lceil k_e \rceil$  would give a minimal value of  $\delta_3$ .

**Group 4: Linear parameters size and logarithmic proof size.** Schemes which correspond to the Group 4 are: Dory [38], Gemini [8], Bulletproofs [11] and Compressed Sigma-Protocol [3]. Taking into account their complexities, we define  $\delta_4$  in the following way:

$$\delta_4(k; c, n) = ack + \frac{nb \cdot \log(ck)}{k}$$

Let  $r = \frac{a}{bnc}$  and  $K = kc$ , then

$$\delta_4(k; c, n) = bnc \left( rK + \frac{\log(K)}{K} \right)$$

One can notice that the optimal value of  $K$  depends on the value of  $r$ : the smaller the value of  $r$ , the larger the value of the optimal  $K$ . We present a plot demonstrating the dependence of the optimal  $K$  on  $r$  in Figure 2 (a). As a result, using this plot, one can choose the ZKP scheme with linear parameters size and logarithmic proof size and infer the optimal value of  $k$  with respect to the parameters of the chosen scheme and parameters of the FL setting.

**Group 5:  $\lceil \log \lceil c \rceil$  parameters size and constant proof size.** Lastly, we describe  $\delta_5$  which corresponds to the Group 5 which consist of only one approach: vnTinyRAM [5].

$$\delta_5(k; c, n) = ack \cdot \log(ck) + \frac{nb}{k}$$

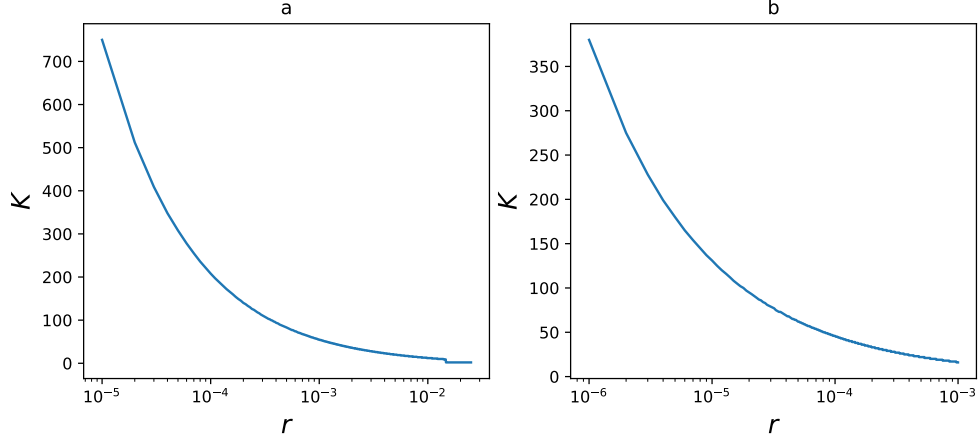


Figure 2: The dependence of the optimal  $K$  on  $r$  for  $\delta_4$  (a) and  $\delta_5$  (b).

Applying the same substitution as for the  $\delta_4$ :

$$\delta_5(k; c, n) = bnc(rK \cdot \log(K) + \frac{1}{K})$$

Similarly, to the previous group, the plot demonstrating the dependence of the optimal  $K$  on  $r$  is presented in Figure 2 (b).

With the above generally applicable strategy, one can determine the optimal number of grouped circuits to prove for various ZKP schemes in order to minimize the storage cost or, in other words, communication cost of a verifiable ZKP-based FL protocol.

Still, once the number of circuits in a group is fixed, the costs get linear in the number of groups. In practice, while the simple strategy discussed above already allows one to handle reasonably large learning tasks, there is a maximal size to a group, e.g., due to memory storage limits. Hence, the asymptotic cost given a bound on memory storage remains linear, and given the ever growing size of ML tasks one could wonder whether one can do even better.

One idea which may address this issue, but has not been tried in practical FL settings, is to compose proofs recursively, in particular to use a ZKP scheme that allows for constructing proofs that attest to the correctness of both circuits and other proofs in the same ZKP scheme. Following this idea, at the beginning a prover creates a proof for the first part of computations and then iteratively generates a proof for the correctness of both the next circuit and all preceding proofs. If successful, this idea then could keep the work of the prover about linear in the number of computations, the memory storage of the prover constant, the proof size and communication cost constant and the verifier cost logarithmic.

Several papers have studied the theory of this type of approach. For example, in the context of incrementally verifiable computations (IVC) [36, 34, 35, 13], proof-carrying data (PCD) [63] and zk Virtual Machine (zkVM) [40]. State-of-the-art solutions mostly rely on a folding scheme, a primitive that reduces the task of checking two instances of some relation to the task of checking a single instance [36]. For instance, authors of Nova [36] were the first to apply the folding scheme to IVC, resulting in a scheme where the prover’s work is linear to the size of the smallest circuit, and the verifier’s time and the proof size are logarithmic. Later, other authors presented advanced versions of the protocol such as SuperNova [34], HyperNova [35], and ProtoStar [13], extending the original approach to more general settings and more efficient arithmetizations.

However, to the best of our knowledge, there are neither works studying an efficient recursive proof composition to verify computations in FL, nor works exploring whether solutions developed for IVC, PCD and zk VM could be adapted to FL. Among all protocols analyzed in Section 3, only in zkFL [56] authors apply a recursive ZKP scheme: Halo 2. Nevertheless, they do not discuss the construction of a minimal circuit or recursion properties of the chosen ZKP and Halo 2 is an early recursive ZKP scheme with linear verifier cost. A linear verifier cost is not practical for large ML tasks. Therefore, the impact of recursion possibilities of the scheme on the FL complexities remains unclear.

We believe that applying more efficient recursive proof composition to verify FL computation remains an open research question and can potentially reduce the gap that prevents the use of ZKPs in practical applications.

## 5 Challenges and future directions

While verifiable cross-silo FL protocols is a well studied area with a plenty of solutions, our analysis reveals several challenges which have not been yet addressed by the research community. Below we discuss potential research gaps and future directions.

Firstly, to the best of our knowledge, there are currently no verifiable FL protocols that fully support verification of both computations performed by clients and server at the same time. Exceptionally, PVD-FL [61] aims to achieve this, however the approach does not guarantee the correctness of all actions of participants and is still vulnerable to poisoning attacks. While some papers, such as [39], consider threats from both the server and clients, authors do not verify the clients' behavior, but analyze the distribution of their values to address potential attacks. Instead, we believe it would be an interesting direction to design a protocol that would enable verification (following the definition in 2.4) of actions of all participants. It would be later intriguing to compare different approaches from the threat models and the efficiency perspectives.

Secondly, we observe that verifiable aggregation is primarily studied for the most popular type of aggregation – averaging of vectors possessed by DOs. Nevertheless, in certain settings, other U-statistics with a kernel of the degree two or larger (e.g. Kendall rank correlation coefficient) could be applied [4], introducing new challenges in the verification process. Based on our analysis, we can see that ZKP-based verification would potentially fit, however finding an efficient solution still remains an open problem.

Thirdly, among the papers considered in this SOK, there are no protocols that are robust against a collusion between client and server to bypass the verification. However, in real world scenarios such collusion might occur. Interestingly, in VerSA [25], the authors do not consider collusion attacks, justifying this choice with the impossibility result shown by Gordon et al. [22]. The authors of the latter paper demonstrated that in specific settings multiclient verifiable computation cannot be achieved in the presence of users colluding with the server. Nonetheless, the specific setting considered does not necessarily apply to FL. For instance, in contrast to the setting from [22], in a verifiable FL protocol clients may exchange messages with each other or prove computations interactively. As a result, this impossibility result does not necessarily show that a verifiable FL protocol cannot be developed to be robust against server-client collusion attacks. We believe that the development of such a protocol is an interesting direction for future research.

Fourthly, we observe that the repetitive nature of FL training is usually overlooked while developing a verifiable protocol. Nevertheless, this property opens up an opportunity to design various optimizations. For instance, in [24], the authors considered this property to combine verification of multiple epochs together, thereby reducing the computational cost of clients. Following similar ideas, in Subsection 4.2, we proposed an optimization based on the observation that instead of repetitive proofs in FL, one can group circuits together before proving, thereby optimizing communication costs. However, we believe that other optimizations, for example, for approaches from other taxonomy groups, require further analysis.

Lastly, as described in Subsection 4.2, to the best of our knowledge, there are no works exploring the applicability of recursive ZKP schemes in the context of FL. In recent years, there has been an active research in the ZKP community to develop such schemes [9, 36, 35]. We anticipate that they deserve a particular attention. Their characteristics may allow for new optimizations and significant reduction in complexities of verifiable protocols.

## 6 Conclusion

In this paper, we presented a systematization of knowledge on verifiable cross-silo FL. We proposed a new taxonomy distinguishing four categories of verification techniques. We described general design patterns for each category and provided an analysis of threat models, computational and communication costs both per client and per server for each protocol. We also discussed how various features of the cross-silo setting impact the verification process and highlighted advantages of ZKP-based protocols. As a continuation of our conclusions, we discussed the applicability of different ZKP schemes for cross-silo FL and optimization strategies to minimize the communication cost. Finally, we described several research challenges revealed in our analysis and indicated future scientific directions.

## 7 Acknowledgments

This project was partially supported by the 'Chair TIP' project funded by ANR, I-SITE, INRIA and MEL, and the Horizon Europe TRUMPET project grant no. 101070038.

## References

- [1] M. Ahmadi and R. Nourmohammadi. zkdf: An efficient and privacy-preserving decentralized federated learning with zero-knowledge proof. <https://synthical.com/article/a8c00457-dadd-4207-bd23-7edaf0188617>, 11 2023.
- [2] A. Arun, C. Ganesh, S. Lokam, T. Mopuri, and S. Sridhar. Dew: A transparent constant-sized polynomial commitment scheme. In *Public-Key Cryptography – PKC 2023: 26th IACR International Conference on Practice and Theory of Public-Key Cryptography, Atlanta, GA, USA, May 7–10, 2023, Proceedings, Part II*, page 542–571, Berlin, Heidelberg, 2023. Springer-Verlag.
- [3] T. Attema and R. Cramer. Compressed  $\sigma$ -protocol theory and practical application to plug & play secure algorithmics. In D. Micciancio and T. Ristenpart, editors, *Advances in Cryptology – CRYPTO 2020*, pages 513–543, Cham, 2020. Springer International Publishing.
- [4] J. Bell, A. Bellet, A. Gascon, and T. Kulkarni. Private protocols for u-statistics in the local model and beyond. In S. Chiappa and R. Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1573–1583, Online, 26–28 Aug 2020. PMLR.
- [5] E. Ben-Sasson, A. Chiesa, E. Tromer, and M. Virza. Succinct Non-Interactive zero knowledge for a von neumann architecture. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 781–796, San Diego, CA, Aug. 2014. USENIX Association.
- [6] M. Blum. Coin flipping by telephone a protocol for solving impossible problems. *SIGACT News*, 15(1):23–27, jan 1983.
- [7] J. Bootle, A. Cerulli, P. Chaidos, J. Groth, and C. Petit. Efficient zero-knowledge arguments for arithmetic circuits in the discrete log setting. In *Proceedings, Part II, of the 35th Annual International Conference on Advances in Cryptology – EUROCRYPT 2016 - Volume 9666*, page 327–357, Berlin, Heidelberg, 2016. Springer-Verlag.
- [8] J. Bootle, A. Chiesa, Y. Hu, and M. Orrù. Gemini: Elastic snarks for diverse environments. In O. Dunkelman and S. Dziembowski, editors, *Advances in Cryptology – EUROCRYPT 2022*, pages 427–457, Cham, 2022. Springer International Publishing.
- [9] S. Bove, J. Grigg, and D. Hopwood. Halo: Recursive proof composition without a trusted setup. *IACR Cryptol. ePrint Arch.*, 2019:1021, 2019.
- [10] C. Brunetta, G. Tsaloli, B. Liang, G. Banegas, and A. Mitrokotsa. Non-interactive, secure verifiable aggregation for decentralized, privacy-preserving learning. *Cryptology ePrint Archive*, Paper 2021/654, 2021.
- [11] B. Bunz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille, and G. Maxwell. Bulletproofs: Short proofs for confidential transactions and more. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 315–334, Los Alamitos, CA, USA, may 2018. IEEE Computer Society.
- [12] B. Bünz, B. Fisch, and A. Szepieniec. Transparent snarks from dark compilers. In A. Canteaut and Y. Ishai, editors, *Advances in Cryptology – EUROCRYPT 2020*, pages 677–706, Cham, 2020. Springer International Publishing.
- [13] B. Bünz and B. Chen. ProtoStar: Generic efficient accumulation/folding for special sound protocols. *Cryptology ePrint Archive*, Paper 2023/620, 2023. <https://eprint.iacr.org/2023/620>.
- [14] C. Fang, Y. Guo, J. Ma, H. Xie, and Y. Wang. A privacy-preserving and verifiable federated learning method based on blockchain. *Computer Communications*, 186:1–11, 2022.
- [15] A. Fiat and A. Shamir. How to prove yourself: Practical solutions to identification and signature problems. In A. M. Odlyzko, editor, *Advances in Cryptology – CRYPTO’ 86*, pages 186–194, Berlin, Heidelberg, 1987. Springer Berlin Heidelberg.
- [16] A. Fu, X. Zhang, N. Xiong, Y. Gao, H. Wang, and J. Zhang. Vfl: A verifiable federated learning with privacy-preserving for big data in industrial iot. *IEEE Transactions on Industrial Informatics*, 18(5):3316–3326, 2022.
- [17] A. Gabizon, Z. J. Williamson, and O.-M. Ciobotaru. Plonk: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. *IACR Cryptol. ePrint Arch.*, 2019:953, 2019.



- [18] H. Gao, N. He, and T. Gao. Sverifl: Successive verifiable federated learning with privacy-preserving. *Information Sciences*, 622:98–114, 2023.
- [19] A. Garg and V. Mago. Role of machine learning in medical research: A survey. *Computer Science Review*, 40:100370, 2021.
- [20] R. Gennaro, C. Gentry, B. Parno, and M. Raykova. Quadratic span programs and succinct nizks without pcps. In T. Johansson and P. Q. Nguyen, editors, *Advances in Cryptology – EUROCRYPT 2013*, pages 626–645, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [21] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989.
- [22] S. D. Gordon, J. Katz, F.-H. Liu, E. Shi, and H.-S. Zhou. Multi-client verifiable computation with stronger security guarantees. In Y. Dodis and J. B. Nielsen, editors, *Theory of Cryptography*, pages 144–168, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.
- [23] J. Groth. On the size of pairing-based non-interactive arguments. In *Proceedings, Part II, of the 35th Annual International Conference on Advances in Cryptology – EUROCRYPT 2016 - Volume 9666*, page 305–326, Berlin, Heidelberg, 2016. Springer-Verlag.
- [24] X. Guo, Z. Liu, J. Li, J. Gao, B. Hou, C. Dong, and T. Baker. Verifl: Communication-efficient and fast verifiable aggregation for federated learning. *IEEE Transactions on Information Forensics and Security*, 16:1736–1751, 2021.
- [25] C. Hahn, H. Kim, M. Kim, and J. Hur. Versa: Verifiable secure aggregation for cross-device federated learning. *IEEE Transactions on Dependable and Secure Computing*, 20(1):36–52, 2023.
- [26] J. Heiss, E. Grunewald, S. Tai, N. Haimerl, and S. Schulte. Advancing blockchain-based federated learning through verifiable off-chain computations. In *2022 IEEE International Conference on Blockchain (Blockchain)*, pages 194–201, Los Alamitos, CA, USA, aug 2022. IEEE Computer Society.
- [27] D. Herhausen, S. F. Bernritter, E. W. Ngai, A. Kumar, and D. Delen. Machine learning in marketing: Recent progress and future research directions. *Journal of Business Research*, 170:114254, 2024.
- [28] C.-F. Hsu, J.-L. Huang, F.-H. Liu, M.-C. Chang, and W.-C. Chen. Fedtrust: Towards building secure robust and trustworthy moderators for federated learning. In *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 318–323, CA, USA, Aug. 2022. IEEE.
- [29] C. Hu, C. Zhang, D. Lei, T. Wu, X. Liu, and L. Zhu. Achieving privacy-preserving and verifiable support vector machine training in the cloud. *IEEE Transactions on Information Forensics and Security*, 18:3476–3491, 2023.
- [30] C. Huang, J. Huang, and X. Liu. Cross-silo federated learning: Challenges and opportunities, 2022.
- [31] C. Huang, J. Wang, H. Chen, S. Si, Z. Huang, and J. Xiao. zkmlaas: a verifiable scheme for machine learning as a service. In *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pages 5475–5480, Rio de Janeiro, Brazil, 2022. IEEE.
- [32] G. Keshavarzkalhori, C. Pérez-Solà, G. Navarro-Arribas, J. Herrera-Joancomartí, and H. Yajam. Federify: A verifiable federated learning scheme based on zk-snarks and blockchain. *IEEE Access*, 12:3240–3255, 2024.
- [33] A. Kosba, D. Papadopoulos, C. Papamanthou, and D. Song. Mirage: Succinct arguments for randomized algorithms with applications to universal zk-snarks. In *29th USENIX Security Symposium*, pages 2129–2146, Boston, MA, USA, 08 2020. USENIX Association.
- [34] A. Kothapalli and S. Setty. SuperNova: Proving universal machine executions without universal circuits. Cryptology ePrint Archive, Paper 2022/1758, 2022. <https://eprint.iacr.org/2022/1758>.
- [35] A. Kothapalli and S. Setty. Hypernova: Recursive arguments for customizable constraint systems. Cryptology ePrint Archive, Paper 2023/573, 2023. <https://eprint.iacr.org/2023/573>.
- [36] A. Kothapalli, S. Setty, and I. Tzialla. Nova: Recursive zero-knowledge arguments from folding schemes. In Y. Dodis and T. Shrimpton, editors, *Advances in Cryptology – CRYPTO 2022*, pages 359–388, Cham, 2022. Springer Nature Switzerland.
- [37] M. Labs. Zero-knowledge proof meets machine learning in verifiability: A survey. <https://drive.google.com/file/d/1tylpowpaqc0hKQtYolP1qvx6R2Gv4IzE/view>, 2023.
- [38] J. Lee. Dory: Efficient, transparent arguments for generalised inner products and polynomial commitments. In K. Nissim and B. Waters, editors, *Theory of Cryptography*, pages 1–34, Cham, 2021. Springer International Publishing.

- [39] Z. Li, H. Yu, T. Zhou, L. Luo, M. Fan, Z. Xu, and G. Sun. Byzantine resistant secure blockchained federated learning at the edge. *IEEE Network*, 35(4):295–301, 2021.
- [40] T. Liu, Z. Zhang, Y. Zhang, W. Hu, and Y. Zhang. Ceno: Non-uniform, segment and parallel zero-knowledge virtual machine. Cryptology ePrint Archive, Paper 2024/387, 2024. <https://eprint.iacr.org/2024/387>.
- [41] F. Luo, S. Al-Kuwari, and Y. Ding. Svfl: Efficient secure aggregation and verification for cross-silo federated learning. *IEEE Transactions on Mobile Computing*, 23(1):850–864, 2024.
- [42] A. Madi, O. Stan, A. Mayoue, A. Grivet-Sébert, C. Gouy-Pailler, and R. Sirdey. A secure federated learning framework using homomorphic encryption and verifiable computing. In *2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS)*, pages 1–8, McMaster University, Hamilton ON, Canada, 2021. IEEE.
- [43] M. Maller, S. Bowe, M. Kohlweiss, and S. Meiklejohn. Sonic: Zero-knowledge snarks from linear-size universal and updatable structured reference strings. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19*, page 2111–2128, New York, NY, USA, 2019. Association for Computing Machinery.
- [44] M. Mohamad, M. Önen, W. Ben Jaballah, and M. Conti. Sok: Secure aggregation based on cryptographic schemes for federated learning. In IACR, editor, *PETS 2023, 23rd Privacy Enhancing Technologies Symposium, 10-15 July 2023, Lausanne, Switzerland (Hybrid Conference)*, pages 140–157, Lausanne, 2023. Rochester NY: Privacy Enhancing Technologies Board. IACR.
- [45] A. Nitulescu. zk-snarks: A gentle introduction. Technical report, Ecole Normale Supérieure, 2019.
- [46] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer. Deep learning for financial applications : A survey. *Applied Soft Computing*, 93:106384, 2020.
- [47] B. Parno, J. Howell, C. Gentry, and M. Raykova. Pinocchio: Nearly practical verifiable computation. In *2013 IEEE Symposium on Security and Privacy*, pages 238–252, San Francisco, California, US, 2013. IEEE.
- [48] A. S. Pinto, A. Abreu, E. Costa, and J. Paiva. How machine learning (ml) is transforming higher education: A systematic literature review. *Journal of Information Systems Engineering and Management*, 8(2):21168, Apr. 2023.
- [49] N. Rodríguez-Barroso, D. Jiménez-López, M. V. Luzón, F. Herrera, and E. Martínez-Cámara. Survey on federated learning threats: Concepts, taxonomy on attacks and defences, experimental study and challenges. *Information Fusion*, 90:148–173, 2023.
- [50] T. Rückel, J. Sedlmeir, and P. Hofmann. Fairness, integrity, and privacy in a scalable blockchain-based federated learning system. *Computer Networks*, 202:108621, 2022.
- [51] C. Sabater, A. Bellet, and J. Ramon. An accurate, scalable and verifiable protocol for federated differentially private averaging. *Machine Learning*, 111, 10 2022.
- [52] I. A. Seres and P. Burcsi. Behemoth: transparent polynomial commitment scheme with constant opening proof size and verifier time. Cryptology ePrint Archive, Paper 2023/670, 2023. <https://eprint.iacr.org/2023/670>.
- [53] A. Tariq, M. A. Serhani, F. Sallabi, T. Qayyum, E. S. Barka, and K. A. Shuaib. Trustworthy federated learning: A survey, 2023.
- [54] G. Tsaloli, B. Liang, C. Brunetta, G. Banegas, and A. Mitrokotsa. Deva: Decentralized, verifiable secure aggregation for privacy-preserving learning. In *Information Security: 24th International Conference, ISC 2021, Virtual Event, November 10–12, 2021, Proceedings*, page 296–319, Berlin, Heidelberg, 2021. Springer-Verlag.
- [55] H. Wang, Z. Zheng, S. Xie, H.-N. Dai, and X. Chen. Blockchain challenges and opportunities: a survey. *International Journal of Web and Grid Services*, 14:352 – 375, 10 2018.
- [56] Z. Wang, N. Dong, J. Sun, W. Knottenbelt, and Y. Guo. zkfl: Zero-knowledge proof-based gradient aggregation for federated learning. *IEEE Transactions on Big Data*, PP(01):1–14, may 2024.
- [57] Z. Xing, Z. Zhang, J. Liu, Z. Zhang, M. Li, L. Zhu, and G. Russello. Zero-knowledge proof meets machine learning in verifiability: A survey, 2023.
- [58] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin. Verifynet: Secure and verifiable federated learning. *IEEE Transactions on Information Forensics and Security*, 15:911–926, 2020.
- [59] X. Zhang, A. Fu, H. Wang, C. Zhou, and Z. Chen. A privacy-preserving and verifiable federated learning scheme. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pages 1–6, Online, 2020. IEEE.
- [60] Y. Zhang and H. Yu. Towards verifiable federated learning. In *International Joint Conference on Artificial Intelligence*, pages 5686–5693, Vienna, Austria, 2022. IJCAI.

- 
- [61] J. Zhao, H. Zhu, F. Wang, R. Lu, Z. Liu, and H. Li. Pvd-fl: A privacy-preserving and verifiable decentralized federated learning framework. *IEEE Transactions on Information Forensics and Security*, 17:2059–2073, 2022.
  - [62] L. Zhao, Q. Wang, C. Wang, Q. Li, C. Shen, and B. Feng. Veriml: Enabling integrity assurances and fair payments for machine learning as a service. *IEEE Transactions on Parallel and Distributed Systems*, 32(10):2524–2540, 2021.
  - [63] Z. Zhou, Z. Zhang, Z. Zhang, and J. Dong. Proof-carrying data from multi-folding schemes. Cryptology ePrint Archive, Paper 2023/1282, 2023.
  - [64] W. Zou, D. Lo, P. S. Kochhar, X.-B. D. Le, X. Xia, Y. Feng, Z. Chen, and B. Xu. Smart contract development: Challenges and opportunities. *IEEE Transactions on Software Engineering*, 47(10):2084–2106, 2021.