

DECENTRALIZED VERIFIABLE CODED STORAGE FOR DISTRIBUTED MACHINE LEARNING

E. Pavlova^{1, *}

¹ Skolkovo Institute of Science and Technology, Moscow, Russia

A storage scheme is proposed for distributed machine learning on NeoFS (or any S3-compatible storage system), providing space efficiency and verifiable data integrity without the need for a consensus layer. Datasets are stored as erasure-coded shards with a single physical replica, and shard headers carry succinct integrity proofs. Clients verify shards by fetching only headers, then reconstruct samples by downloading any k shards in parallel. NeoFS storage policies handle placement, while an optional on-chain manifest anchors dataset commitments for auditability. We outline the shard format, header-only verification, and the end-to-end sampling pipeline. A LeNet-5 pilot shows that verification and decoding do not limit training throughput.

Схема хранения для распределенного машинного обучения на базе NeoFS обеспечивает компактность данных и верифицируемую целостность без необходимости консенсуса. Наборы данных кодируются стирающими кодами и хранятся в виде шардов с единственной физической репликой, заголовки которых содержат доказательства целостности. Проверка выполняется по заголовкам, восстановление — по любым k шардам. Политики NeoFS управляют размещением, а ончейн-манифест фиксирует обязательства для аудита. Эксперимент на модели LeNet-5 подтверждает, что проверка и декодирование не снижают производительность обучения.

PACS: 89.70.—a; 07.05.Bx; 02.70.—c; 89.20.Ff

* E-mail: ekaterina.pavlova@skoltech.ru