

 [huggingface](#) / [lerobot](#) Public[Code](#) [Issues](#) 343 [Pull requests](#) 305 [Actions](#) [Security and quality](#) [Ins](#)

Fix: Replace unsafe pickle with safetensors + JSON in async inference #3048



Open

Chocapikk wants to merge 4 commits into [huggingface:main](#) from[Chocapikk:fix/replace-pickle-with...](#) Conversation 0Commits 4Checks 1Files changed 4Chocapikk commented [on Feb 27](#) • edited ▾

Summary

Fixes [#3047](#)

Replaces all `pickle.loads()` / `pickle.dumps()` calls in the async inference pipeline with safe alternatives:

- **safetensors** for tensor data (`torch.Tensor`, numpy arrays)
- **JSON** for scalar metadata and configuration

This eliminates the CWE-502 (Deserialization of Untrusted Data) vulnerability that allowed unauthenticated Remote Code Execution via crafted pickle payloads over the gRPC channel.

Changes

- **New:** `safe_serialization.py` - Serialization module using a simple wire format: `[4-byte JSON length][JSON metadata][safetensors tensor data]`
 - `serialize/deserialize_policy_config` - JSON-only (no tensors needed), handles both `PolicyFeature` dataclass instances and plain dicts
 - `serialize/deserialize_observation` - Separates tensor data (safetensors) from scalar metadata (JSON), preserves numpy vs torch types
 - `serialize/deserialize_actions` - Stores action tensors in safetensors with JSON metadata for timestamps/timesteps
- **Modified:** `policy_server.py` - Replaced 3 pickle calls:

- `SendPolicyInstructions` : `pickle.loads()` -> `deserialize_policy_config()`
 - `SendObservations` : `pickle.loads()` -> `deserialize_observation()`
 - `GetActions` : `pickle.dumps()` -> `serialize_actions()`
- **Modified:** `robot_client.py` - Replaced 3 pickle calls:
 - `start()` : `pickle.dumps()` -> `serialize_policy_config()`
 - `send_observation()` : `pickle.dumps()` -> `serialize_observation()`
 - `receive_actions()` : `pickle.loads()` -> `deserialize_actions()`
 - **New:** `test_safe_serialization.py` - Roundtrip tests for all three data types + type validation + security regression test

Design Decisions

- Uses `safetensors` which is already a project dependency and was built by Hugging Face specifically as a safe alternative to pickle for ML tensors
- Preserves numpy vs torch tensor types through metadata flags
- No changes to the protobuf definitions or gRPC service interface - the `bytes` fields are reused as-is
- Type-checking on deserialization prevents cross-type confusion

Test Results

All tests pass:



```
tests/async_inference/test_safe_serialization.py::TestPolicyConfigSerialization::test_roundtrip PASSED
tests/async_inference/test_safe_serialization.py::TestPolicyConfigSerialization::test_rejects_invalid PASSED
tests/async_inference/test_safe_serialization.py::TestPolicyConfigSerialization::test_no_arbitrary PASSED
tests/async_inference/test_safe_serialization.py::TestObservationSerialization::test_roundtrip PASSED
tests/async_inference/test_safe_serialization.py::TestObservationSerialization::test_roundtrip PASSED
tests/async_inference/test_safe_serialization.py::TestObservationSerialization::test_rejects_invalid PASSED
tests/async_inference/test_safe_serialization.py::TestActionsSerialization::test_roundtrip PASSED
tests/async_inference/test_safe_serialization.py::TestActionsSerialization::test_empty_action PASSED
tests/async_inference/test_safe_serialization.py::TestActionsSerialization::test_rejects_invalid PASSED
tests/async_inference/test_e2e.py::test_async_inference_e2e PASSED


10 passed in 5.40s
```

Breaking Change

This is a **wire format change** - clients and servers must both be updated together. Since the async inference feature is experimental (as noted by maintainers), this should be acceptable.

  [Fix: Replace unsafe pickle serialization with safetensors + JSON in a...](#)   [7415a56](#)

  [github-actions](#) [Bot](#) added the [tests](#) label [on Feb 27](#)

  [Chocapikk](#) changed the title [Fix: Replace unsafe pickle with safetensors + JSON in async inference](#) [WIP: Fix: Replace unsafe pickle with safetensors + JSON in async inference](#) [on Feb 27](#)

  [Chocapikk](#) marked this pull request as draft [2 months ago](#)

  [Chocapikk](#) mentioned this pull request [on Feb 27](#)

[Security: Unsafe pickle deserialization in async inference enables Remote Code Execution \(CWE-502\) #3047](#)

[Open](#)

  [Fix: Handle plain dict lerobot_features in safe serialization](#)   [edbf384](#)

  [Chocapikk](#) marked this pull request as ready for review [2 months ago](#)

  [Chocapikk](#) changed the title [WIP: Fix: Replace unsafe pickle with safetensors + JSON in async inference](#) [Fix: Replace unsafe pickle with safetensors + JSON in async inference](#) [on Feb 27](#)

 [Chocapikk](#) added 2 commits [2 months ago](#)

  [Refactor: Clean up comments and test naming](#)  [27f2a66](#)

  [Refactor: DRY type validation into _unpack and clean up tests](#)  [55746a6](#)

  [imstevenpmwork](#) mentioned this pull request [3 weeks ago](#)

Release 0.6.0 #3134

Open

Sign up for free to join this conversation on GitHub. Already have an account? [Sign in to comment](#)

Reviewers

No reviews

Assignees

No one assigned

Labels

tests

Projects

None yet

Milestone

No milestone

Development

Successfully merging this pull request may close these issues.

Security: Unsafe pickle deserialization in async inference enables Remote Code Execution (CWE-502)

1 participant

