SDL в искусственном интеллекте

Андрей Федотов

23 сентября 2022

ИСП РАН

Непрерывный анализ фреймворков ИИ

Кто занимается этим в мире?

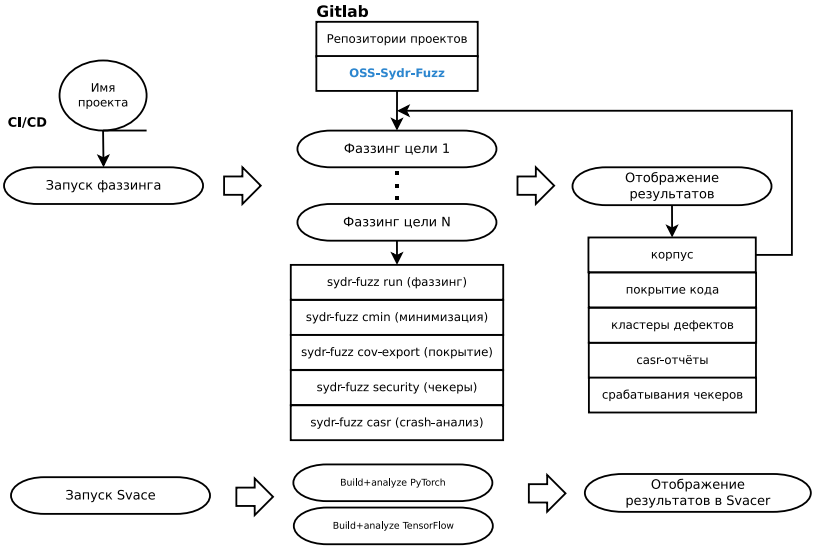
- Проект OSS-Fuzz от Google производит фаззинг фреймворка TensorFlow. Наш патч github.com/google/oss-fuzz/pull/7704 возродил фаззинг TensorFlow в OSS-Fuzz.
- Фреймворк PyTorch не представлен в OSS-Fuzz.

Уязвимости постоянно появляются: *The hidden dangers of loading open-source AI models* (youtube.com/watch?v=2ethDz9KnLk)

Наша цель: непрерывный статический и динамический анализ фреймворков ИИ и сопутствующих проектов с помощью Svace и Sydr.

Система непрерывного анализа ИИ

ИСП РАН



Найденные ошибки в фреймворках ИИ

ИСП РАН

Проект	Ошибки	Исправлений	Принято в upstream
TensorFlow	1	1	1
PyTorch	7	7	6
Torchvsion	1	1	1
OpenCV	1	-	-
miniz	1	1	1

Проект	CRITICAL	Confirmed	Won't fix
TensorFlow	535	38	87
PyTorch	122	41	18

Срабатывания Svace

Подробности о найденных ошибках тут: github.com/ispras/oss-sydr-fuzz

Планы и направления развития

- Разработка web-интерфейса для результатов фаззинга;
- Подготовка патчей в TensorFlow и PyTorch по результатам Svace;
- Тестирование TensorFlow другими фаззерами: Sydr&AFL++;
- Добавление новых фаззинг целей и проектов.