Разработка доверенных версий фреймворков машинного обучения

Андрей Федотов infosec.exchange/@anfedotoff

2 декабря 2022

ИСП РАН

Непрерывный анализ фреймворков ИИ

Кто занимается этим в мире?

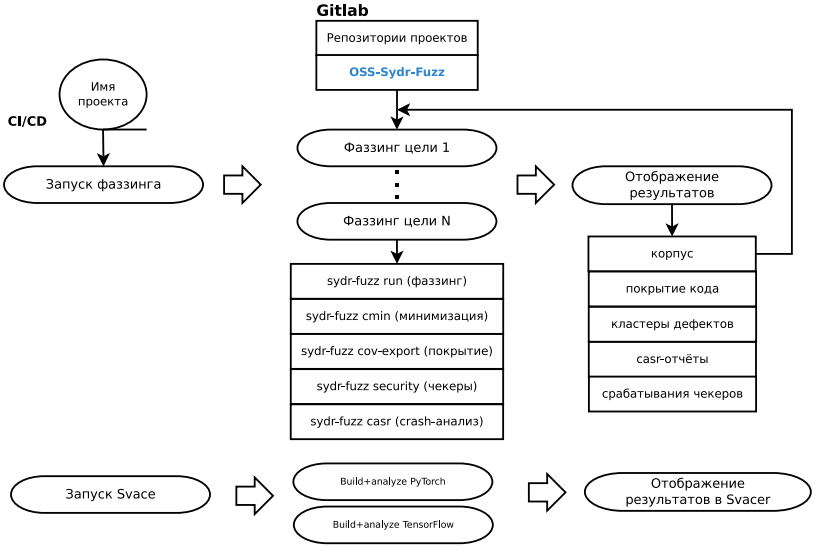
- Проект OSS-Fuzz от Google производит фаззинг фреймворка TensorFlow. Наш патч github.com/google/oss-fuzz/pull/7704 возродил фаззинг TensorFlow в OSS-Fuzz.
- Фреймворк PyTorch не представлен в OSS-Fuzz.

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 (Sydr|Svace)	Исправлений	Принято в upstream
TensorFlow	16 (1|15)	16	1
PyTorch	20 (7|13)	20	19

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

Символьные предикаты безопасности

- предикаты нацелены на поиск ошибок выхода за границы массива, целочисленного переполнения, деления на ноль;
- предикаты применяются после фаззинга на минимизированном корпусе с хорошим покрытием;
- срабатывания верифицируются на санитайзерах (asan+ubsan);
- уникализация срабатываний;
- предикаты показывают 96% точность обнаружения ошибок предикатами безопасности на наборе тестов Juliet.

Интересная ошибка в зависимостях OpenCV

ИСП РАН

github.com/opencv/opencv/issues/22284

opencv/3rdparty/openjpeg/openjp2/image.c:134:

```
l_y1 = p_cp->ty0 + (p_cp->th - 1U) * p_cp->tdy; /* can't overflow */
```

Can't overflow? But we can!

Срабатыване символьного чекера Sydr:

opj_image_comp_header_update:/opencv/3rdparty/openjpeg/openjp2/image.c:134 - imul r15d, eax - **unsigned integer overflow**

Автоматическое подтверждение дефекта санитайзерами:

/opencv/3rdparty/openjpeg/openjp2/image.c:134:40: runtime error: **unsigned integer overflow**: 2 * 4278190076 cannot be represented in type 'unsigned int'

TensorFlow Бесконечный цикл в wav_io.cc

ИСП РАН

tensorflow/pull/56455

```
235   std::string found_text;                                    bext
236   TF_RETURN_IF_ERROR(ReadString(wav_string, 4, &found_text, &offset));
237   while (found_text != kFormatChunkId) {
238     // Padding chunk may occur between "WAVE" and "fmt ".
239     // Skip JUNK/bext/etc field to support for WAV file with either JUNK Chunk,
240     // or broadcast WAV where additional tags might appear.
241     // Reference: the implementation of tfio in audio_video_wav_kernels.cc,
242     //            https://www.daubnet.com/en/file-format-riff,
243     //            https://en.wikipedia.org/wiki/Broadcast_Wave_Format
244     if (found_text != "JUNK" && found_text != "bext" && found_text != "iXML" &&
245         found_text != "qlty" && found_text != "mext" && found_text != "levl" &&
246         found_text != "link" && found_text != "axml") {
247       return errors::InvalidArgument("Unexpected field ", found_text);
248     }
249     uint32 size_of_chunk;
250     TF_RETURN_IF_ERROR(ReadValue<uint32>(wav_string, &size_of_chunk, &offset));
251     TF_RETURN_IF_ERROR(                                      -8
252         IncrementOffset(offset, size_of_chunk, wav_string.size(), &offset));
253     TF_RETURN_IF_ERROR(ReadString(wav_string, 4, &found_text, &offset));
254   }
```

Исправление ошибки в wav_io.cc

tensorflow/pull/56455

```
86  86    // Handles moving the data index forward, validating the arguments, and avoiding
87  87    // overflow or underflow.
88      - Status IncrementOffset(int old_offset, size_t increment, size_t max_size,
    88    + Status IncrementOffset(int old_offset, int increment, size_t max_size,
89  89                             int* new_offset) {
90  90      if (old_offset < 0) {
91  91        return errors::InvalidArgument("Negative offsets are not allowed: ",
92  92                                       old_offset);
93  93      }
    94    +   if (increment < 0) {
    95    +     return errors::InvalidArgument("Negative increment is not allowed: ",
    96    +                                    increment);
    97    +   }
94  98      if (old_offset > max_size) {
95  99        return errors::InvalidArgument("Initial offset is outside data range: ",
96 100                                       old_offset);
```

Планы и направления развития

ИСП РАН

- Добавление новых фаззинг целей для C++ компонентов.
- Фаззинг Python частей фреймворков.
- Дальнейшая разметка предупреждений от Svace.
- Исправление найденных в будущем дефектов.