Исследование и разработка методов динамического анализа для определения входных данных влияющих на выполнение условных переходов

Дьячков Л.А.

Руководитель: к.фм.н, с.н.с. Курмангалеев Ш.Ф.

5 Апреля 2019

ИСП РАН

Глобальные задачи

Разработка инструмента, позволяющего определять какие байты во входных файлах влияют на условные переходы в программе.

- Изучение инструментов динамического анализа помеченных данных и динамического символьного выполнения
- Разработка инструментов для тестирования инструментов и оценки качества их работы
- Доработка выбранного инструмента.

Решаемые подзадачи | OSS Fuzz

Задача Для сравнения инструментов использовать существующию инфраструктуру Google OSS-fuzz

Результаты

- Изучена инфраструктура google os fuzz.
- Получено представление о работе afl и libfuzz (в меньшей степени) с точки зрения пользователя
- На jenkins заведен job, запускающий afl фаззер при помощи oss fuzz
- Понимание как адаптировать имеющуюся инфраструктуру для оценки инструментов taint анализа не получено

Решаемые подзадачи | Разработка библиотеки для снятия и анализа трасс

Задача. Для сравнения инструментов необходимо разработать вспомогательную библиотеку, решающую следующие задачи:

- Сбор информации об условных переходах (адрес, опкод, был ли совершен переход, является ли инструкция помеченной)
- Подсчет различных метрик (длина трассы, количество уникальных прыжков, количество помеченных прыжков)
- Возможность интеграции в проекты на языках C и python

Решаемые подзадачи | Разработка библиотеки для снятия и анализа трасс

Результаты:

- Разработана и покрыта тестами библиотека *insrumentation-lib* на языке программирование Rust, предоставляющая интерфейсы для сбора трассы и подсчета метрик, а также поддерживающая сериализацию и десериализацию трассы.
- Реализовна сборка в виде динамической библиотеки, статической библиотеки и в виде библиотеки для языков python2 и python3
- По техническим (PinCRT) причинам оказалось невозможно использовать библиотеку как предполагалось в проектах, использующих pin3.

Задача. Изучить реализацию и методы работы существующих инструментов динамического анализа, интегрировать в них разработанную библиотеку

- triton
- libdft
- taintgrind
- moflow

Решаемые подзадачи | Изучение инструментов

Результаты. Изучить реализацию и методы работы существующих инструментов динамического анализа, интегрировать в них разработанную библиотеку

- Для triton все реализовано в соответсвии с планом.
- Для libdft часть работ была проделана А. Харченко, из-за PinCRT было принято решение реализовать снятие интересующей информации без помощи *instrumentation-lib*. Использовать утилиту на python на основе *instrumentation-lib* для парсинга.
- Для taintgrind все реализовано в соответсвии с планом.
- Для moflow работа проделана Шамилем. (Но т.к. moflow в итоге был выбран, его изучение было проведено позднее)

Задача. Сравнить работу инструментов на тестовых примерах из набора LAVA.

Результаты.

- Были разработаны скрипты для
  - параллельного запуска инструмента на тестовых примерах.
  - генерирование таблицы с результатами метрик.
- на сервере ibis были проведены запуски инструментов и снятиы интересующие результаты.

Решаемые подзадачи | Генерирование тестовых программ для оценки качества taint

Задача. Разработать генератор C программ, состоящих из последовательности вложенных if выражений над элеменатими последовательности де Брёйна.

Результаты: Успешно генерируются программы вида:

```
int main(int argc, char** argv)
{
  ... // this is one-byte examle, 4-bytes generates as well
  if (( data[ 0 ] | 110 ) < 143) {
   data[1] ^= 56;
   counter++;
   if (( data[ 1 ] ^ 192 ) != 50) {
    data[2] ^= 175;
    counter++;
   }
  }
  printf("%d", counter);
}
```

Все выражения внутри if оказываются истины, так как в противном случае условие заменяется на обратное в процессе генерации.

Задача. Улучшить механизм распространение пометок в Moflow.
Проблемы:

- Нет поддержки нескольких тегов для области памяти/регистра. Специальный тег *MIXED_TAINT* для случая нескольких тегов. **(Решено)**
- Отсутсвие учета семантики инструкции при распространении пометок
- Поддерживаются только старшие регистры. **(В работе, планируется закончить до конца следующей недели)**