# Разбиение дизассемблированного кода на линейные участки

АНИСИМОВ КОНСТАНТИН 444ГР
НАУЧНЫЙ РУКОВОДИТЕЛЬ
БАКЛАНОВСКИЙ М.В.

# Линейный участок

- Структурная единица кода
- Часть кода, которая всегда исполняется с начала до конца
- Только в конце возможна передача управления

# Зачем?

Построение графа потока управления

- Статическая бинарная трансляция
- Реверс-Инжиниринг программ
- Обфускация программы
- Автоматическое распараллеливание кода
- Многоуровневая архитектура кода(МАК)

# Постановка задачи

- Разбиение скомпилированного кода на линейные участки
- Поиск эвристик для разбиения произвольного (обфусцированного) кода на линейные участки
- Сбор статистики о бинарных файлах
  - Как часто встречаются разные непростые ситуации
  - Прочая статистика о исполняемых файлах

- Изначально дан только исполняемый файл

# Трудности

- Не всегда возможность статически разбить код на линейные участки полностью существует
- Не всегда возможно узнать, что мы дизассемблировали код полностью
  - Недостижимый(мертвый код) код - до 90 %

# Трудности

- Подмена адреса возврата функции
- Костенная передача управления
  - Адресу из памяти
  - Адресу из регистра
  - Адресу из памяти по регистру
- API  CallBacks (CreateThread etc)

# Косвенная передача управления

Может возникнуть в процессе компиляции как результат:

- Виртуальных функций
- Таблицы функций
- Callback’и

# Исходные данные

- Точки входа в исполняемом файле:
  - EntryPoint
  - ExportTable
  - TLS Callback
- Секция .pdata

# Результаты

- Создан инструмент для разбиения на линейные участки, сбора статистики, и сравнения эвристик
- Статистика на 1500 x64 исполняемых файлов (exe,dll) ОС Windows
  - 352 Mb исполняемого кода
- Добавлено две эвристики
- Собрана статистика для анализа эвристик
  - Было 15% в среднем стало 87 % от исполняемой секции файла

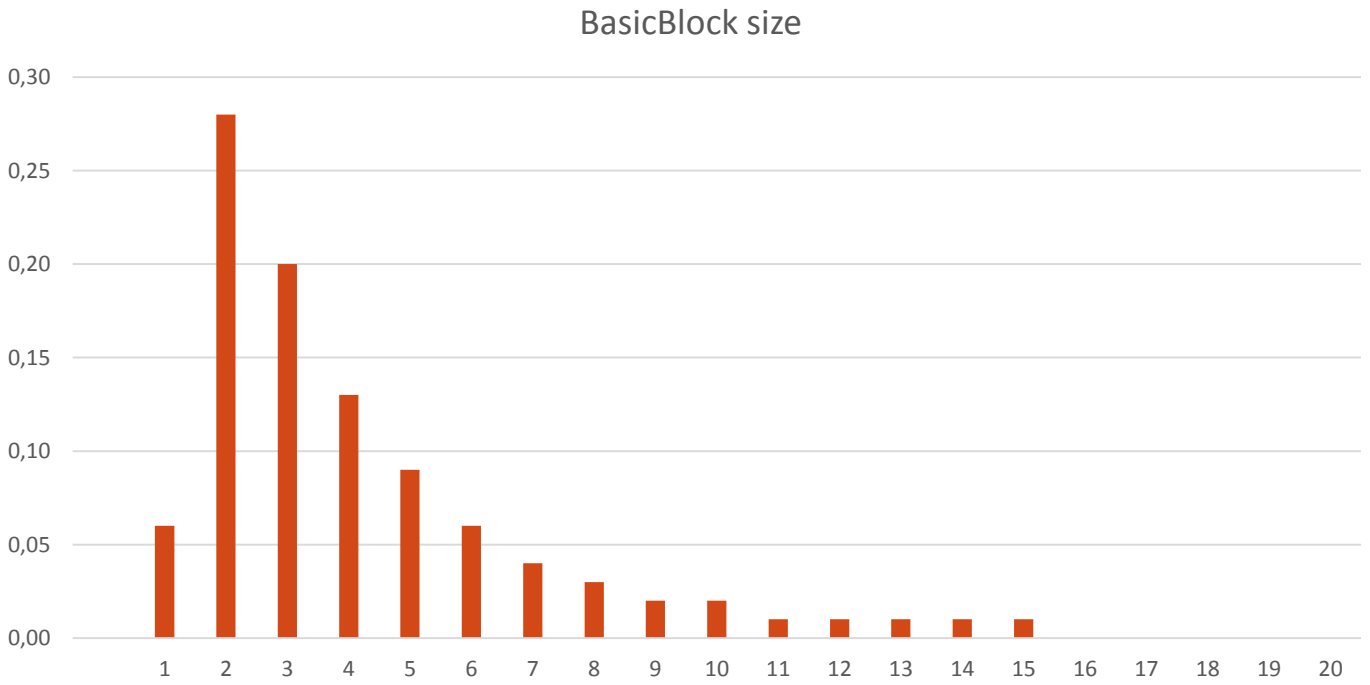
# Статистика

Размер линейного участка:

average : 4.29

2% minimum : 1

2% maximum : 18



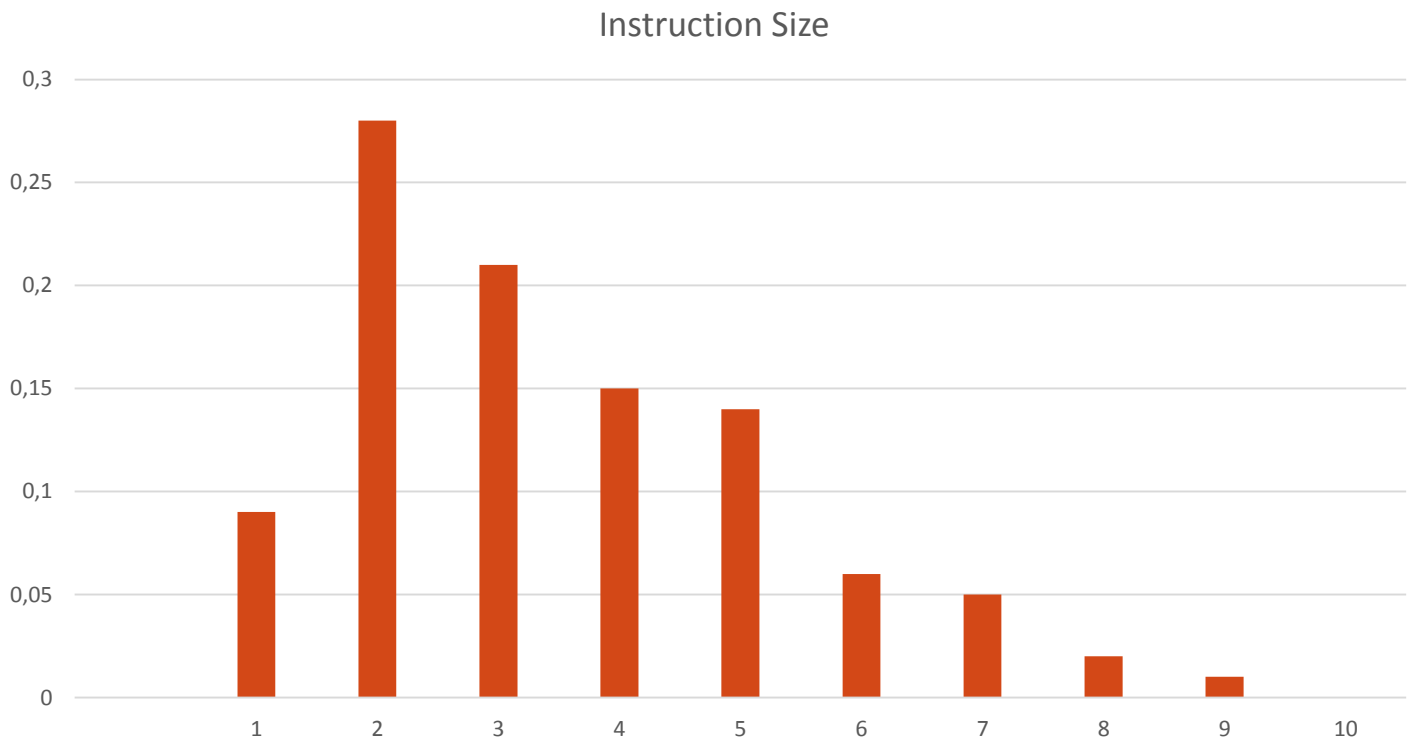
# Статистика

Размер инструкции:

average : 3.9

2% minimum : 1

2% maximum : 8



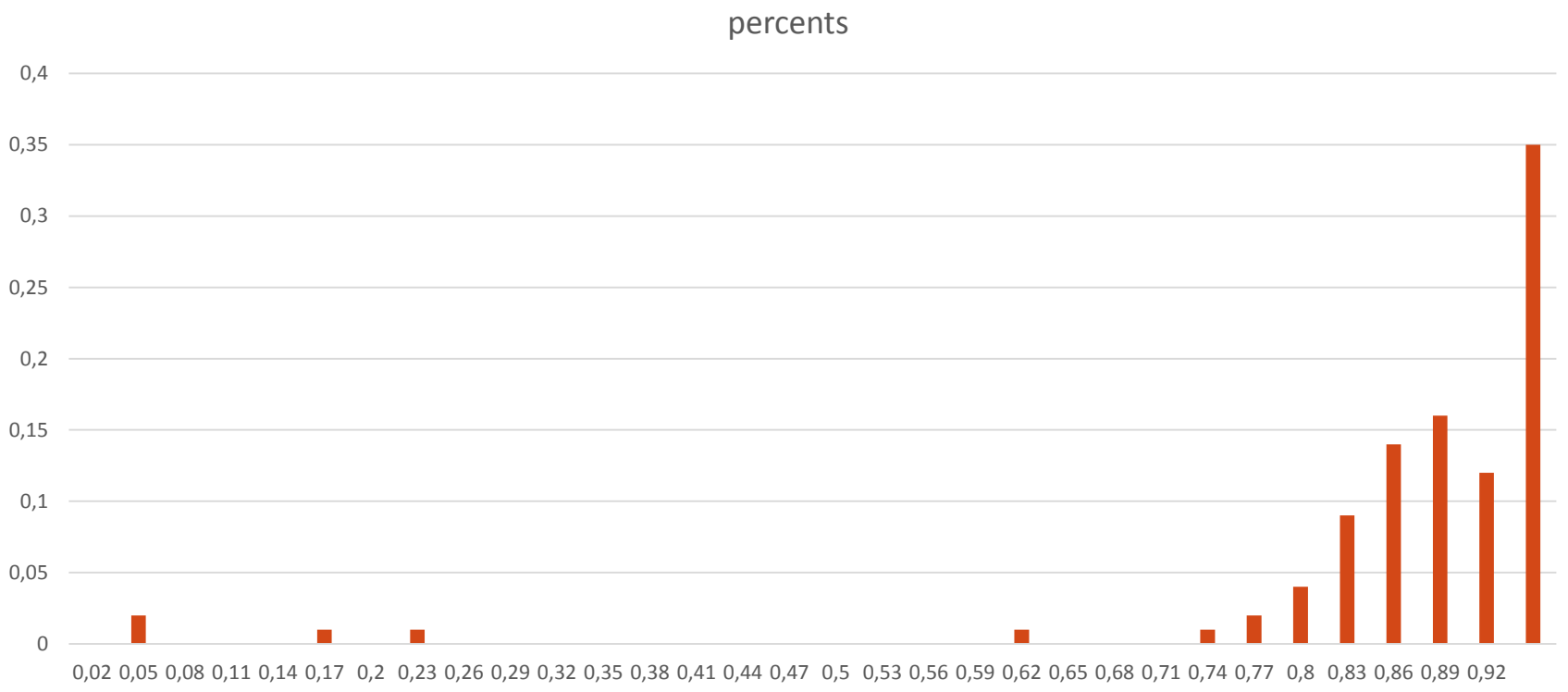
# Статистика

Доля разбиения на линейные участки, со всеми эвристиками:

average : 0.85

2% minimum : 0.22

2% maximum : 0.95



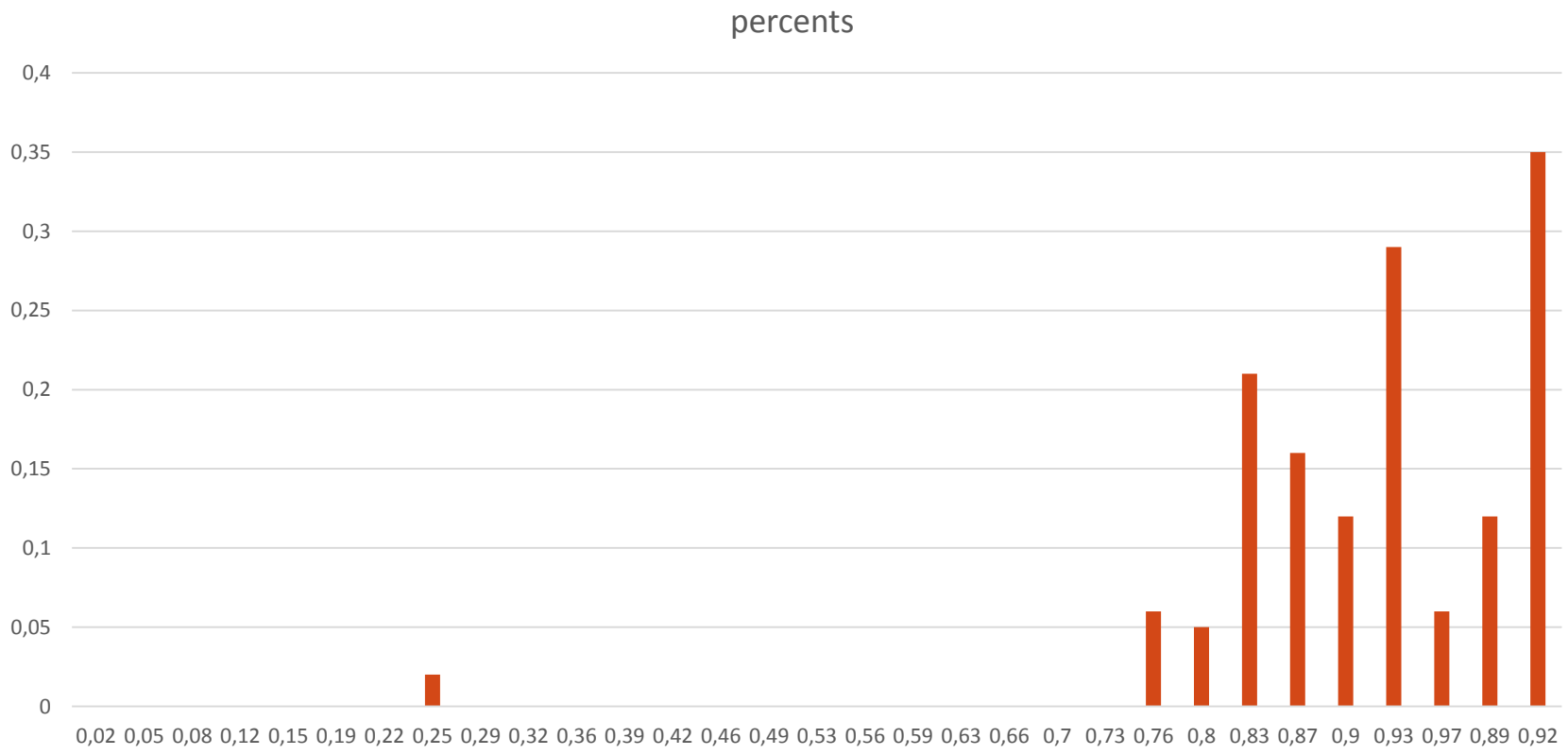
# Статистика

Доля разбиения на линейные участки, только данные для обработки исключений:

average : 0.80

2% minimum : 0.70

2% maximum : 0.92



# Статистика

Доля разбиения на линейные участки, анализируя операнды памяти:

average : 0.36

2% minimum : 0.025

2% maximum : 0.88