О некоторых ограничениях полносистемного анализа помеченных данных

М.А. Климушенкова, М.Г. Бакулин, В.А. Падарян,

П.М. Довгалюк, Н.И. Фурсова, И.А. Васильев

Анализ помеченных данных

Анализ помеченных данных - метод исследования программы путём внесения пометок и отслеживания их распространения по потокам данных

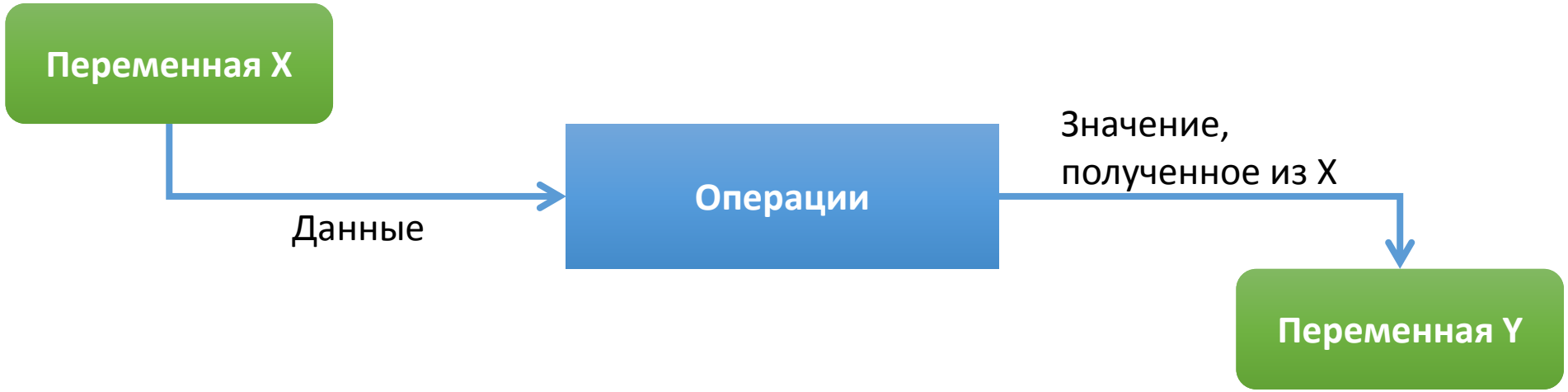
- Отслеживание данных, полученных из недоверенных источников. Если они оказывают непосредственное влияние на счетчик инструкций и исполняемый код, то такие данные следует признать небезопасными

- Отслеживание чувствительных данных.
  Обнаружение «утечек» данных или их несанкционированного использования

Анализ помеченных данных
Пометки

Если источник значения переменной X является недоверенным, то переменная X помечается

Анализ помеченных данных
Поток данных



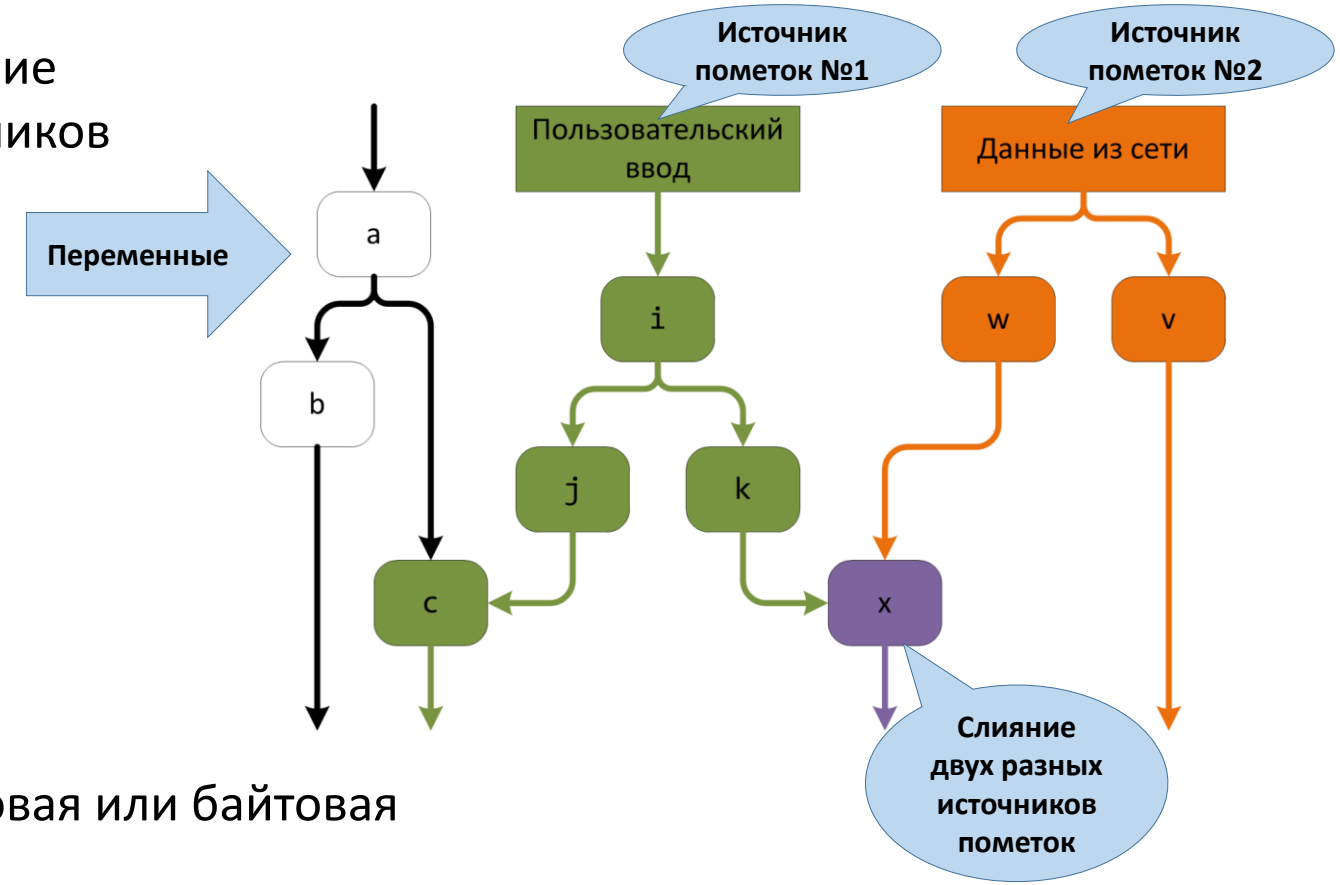
Операция или последовательность операций, которые использует значение переменной X, чтобы получить значение для переменной Y, порождает поток от X к Y

Анализ помеченных данных
Поток данных

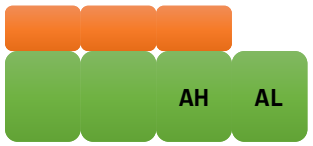
Если операция использует помеченное значение переменной X, чтобы получить значение для переменной Y, то значение X становится помеченным

Распространение меток

Одновременное продвижение пометок от разных источников



Гранулярность пометок: битовая или байтовая



EAX

EDX

Анализ процесса vs анализ системы

- Анализ отдельного процесса
  - Метки не распространяются через код ядра
  - Необходим дополнительный анализ библиотечных функций
  - libdft, Dytan, Minemu
- Полносистемный анализ
  - Существенное замедление
  - TaintBochs, Timescope, DECAF, ARGOS, PANDA

Известные работы

1. PANDA, 2013, Georgia Tech, MIT Lincoln Laboratory
   - Инструмент основан на QEMU 1.0.1
   - Автоматическая трансляция вспомогательных функций и TCG-кода в байткод LLVM, на уровне которого происходит анализ
   - Низкая скорость работы
2. DECAF, 2014, Syracuse University
   - Основан на QEMU 1.0 и BitBlaze
   - Пометки распространяются при помощи инструментального TCG-кода и модифицированных вспомогательных функций
3. ARGOS, 2008, Vrije Universiteit Amsterdam
   - Последняя версия портирована на QEMU 1.1
   - Реализован только для архитектуры i386
   - Предназначен для использования как т.н. honeypot, все входящие сетевые данные автоматически помечаются

Реализация полносистемного анализа

- Где хранить пометки?
- Когда добавлять пометки?
- Как распространять пометки?
- Как обнаружить недопустимую ситуацию?

Реализация полносистемного анализа в QEMU

- Где хранить пометки?
  - Теневая память
  - Теневые регистры
- Когда добавлять пометки?
  - Пометка сетевого трафика
  - Пометка областей памяти или регистров
- Как распространять пометки?
  - Добавление инструкций, продвигающих пометки в промежуточное TCG представление. Инструкции добавляются если пометок нет
  - Изменение вспомогательных функций
- Как обнаружить недопустимую ситуацию?
  - Помеченный счетчик инструкций

Недостаточная помеченность

- Помечено меньше данных, чем необходимо согласно логике работы кода
- Причины:
  - Адресные зависимости

    add esi, eax

    mov [esi], 0x2A
  - Зависимости по управлению

```
if (x == 42)
    y = 0;
else
    y = 1;
```

Избыточная помеченность

- Помечено больше данных, чем необходимо согласно логике работы кода
- Причины:
  - Операции с константным результатом (XOR)
  - Односторонние функции (хэш)
  - Адресные зависимости

```
add esi, eax
cmp esi, 6
ja LN7
jmp LN11[esi*4]
```

Особенности реализации

- Блочная память
- Регистры устройств
- DMA транзакции
- USB-устройства
- Распространение пометок через инструкции FPU, MMX, SSE

DECAF: Ложноположительные результаты

Причины:

- Отсутствие обработки DMA транзакций
- Отсутствие корректной работы с синхронными, асинхронными прерываниями и системными вызовами

Результат:

- Количество ложноположительных срабатываний снизилось с 80% до 25%

Трудности полносистемного анализа

- Теоретические
  - Адресные зависимости
  - Зависимости по управлению
  - Константные и односторонние функции

- Практические
  - Недетерминированная работа симулятора
  - Работа с устройствами
  - DMA-транзакции
  - Обработка прерываний
  - Поддержка FPU, MMX, SSE