НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Факультет компьютерных наук
Курсовой проект
Программный модуль для выделения логически связанных потоков в высокоскоростном сетевом трафике

Работу выполнил студент 3 курса группы БПИ-197

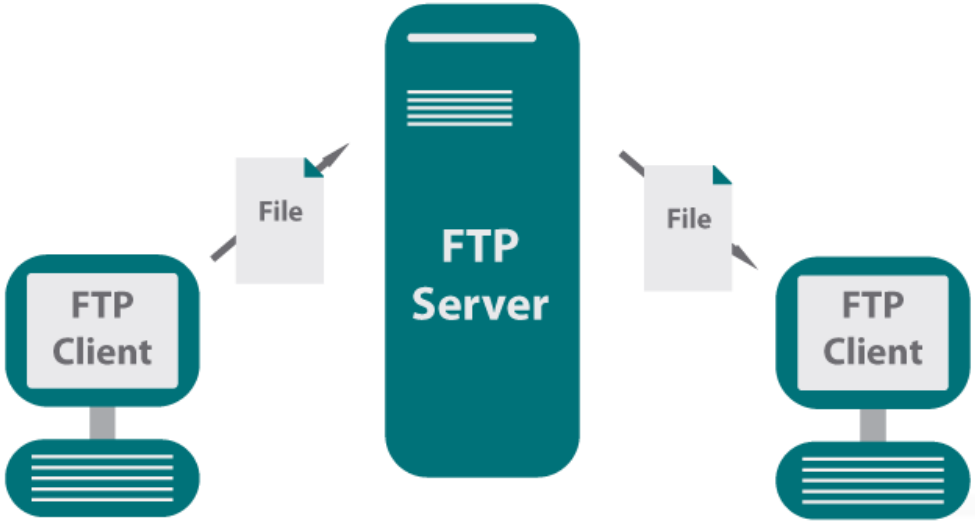
Глущенко Захар Сергеевич

Научный руководитель

Старший научный сотрудник ИСП РАН, к.ф.-м.н.

Гетьман Александр Игоревич

ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ



В современном мире, где все завязано на передачи данных через интернет становиться все более важным и востребованным анализировать передаваемый трафик по множеству причин, начиная от проверки пропускной способности и балансировки нагрузки и заканчивая борьбой с преступностью и пиратством в частности.

Данная работа посвящена анализу трафика использующего протокол FTP. Данный протокол появился в 1971 году и является одним из старейших прикладных протоколов, но из-за своей простоты и удобства в использовании остается популярным и по сей день. Построен на архитектуре «клиент-сервер» и использует разные сетевые соединения для передачи команд и данных между клиентом и сервером.

ОСНОВНЫЕ ПОНЯТИЯ, ОПРЕДЕЛЕНИЯ, ТЕРМИНЫ

5-tuple – Кортеж, состоящий из 5 элементов:

- IP адрес отправителя
- Номер port отправителя
- IP адрес получателя
- Номер port получателя
- Протокол (TCP | UDP)

Регулярные выражения – используемый в компьютерных программах, работающих с текстом, формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. Для поиска используется строка-образец (паттерн), состоящая из символов и метасимволов и задающая правило поиска.

Конфигурационный файл - файл, содержащий необходимые настройки для работы программы, которые можно редактировать для изменения поведения исходной программы.

АКТУАЛЬНОСТЬ РАБОТЫ

Данный программный модуль может быть использован в составе системы анализа трафика, разрабатываемой в Институте системного программирования РАН и применяться, когда необходимо найти логически связанные потоки и объединить их.

Система в состав которой включен данный модуль может быть использована для помощи в обнаружении различных нарушений и сборе доказательств этих нарушений.

ЦЕЛЬ И ЗАДАЧИ РАБОТЫ

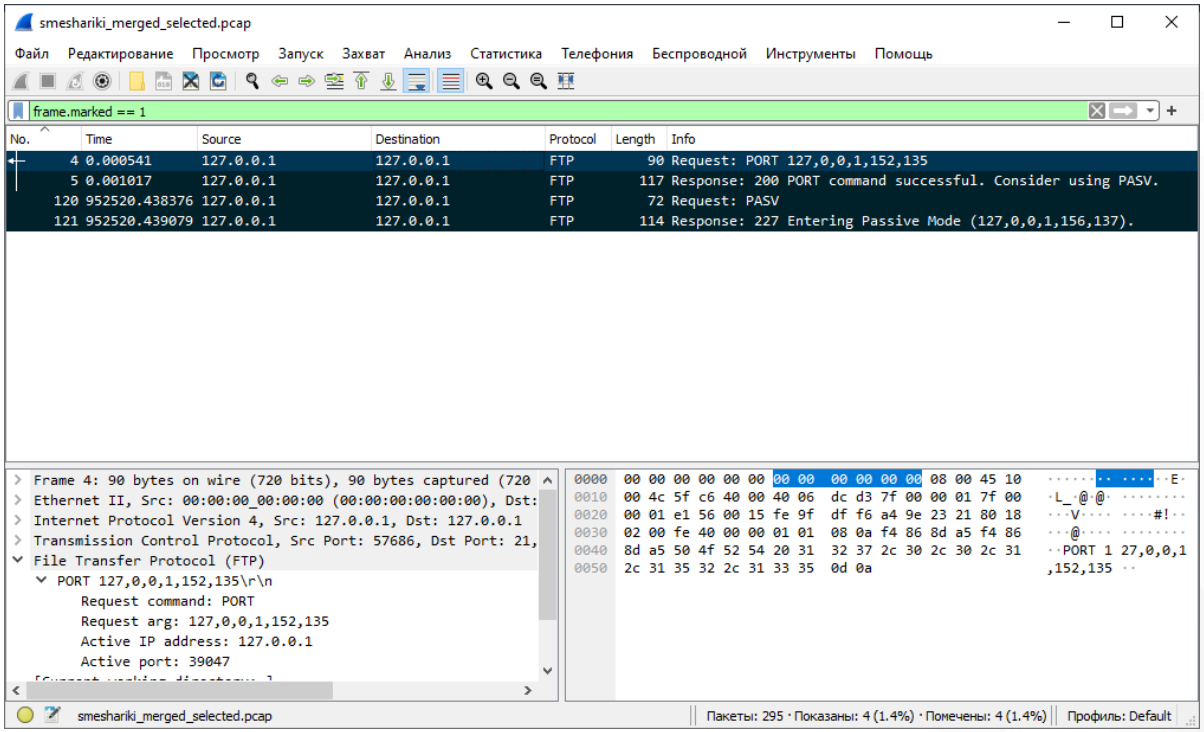
Цель работы

Выделение логически связанных потоков в сетевом трафике.

Задачи работы

1. Обнаружить пакет с командой содержащей адрес нового соединения данных
2. Извлечь информацию и направить ее в соответствующие модули
3. При получении полной информации о потоках, отобрать и сохранить их.
4. Объединить сохраненные потоки управляющего соединения с созданными им соединениями данных

ФИЛЬТРАЦИЯ ПАКЕТОВ



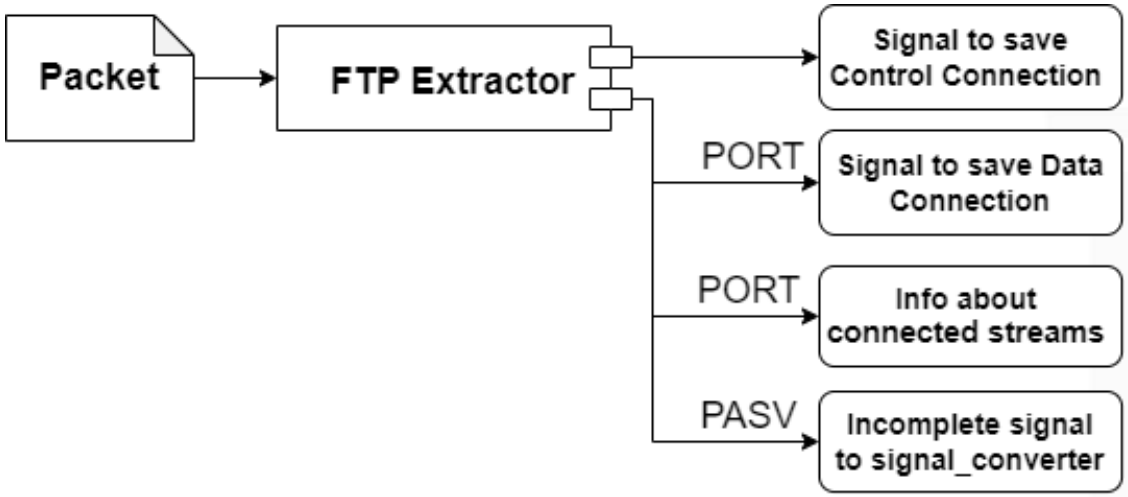
Команда содержащая адрес нового соединения передается в управляющем соединении.

Возможны два способа установления соединения данных:

- Активное
- Пассивное.

Для фильтрации пакетов используется следующее регулярное выражение:
(227 Entering Passive Mode)|(PORT ([0-9]{1,3},){5}[0-9]{1,3})

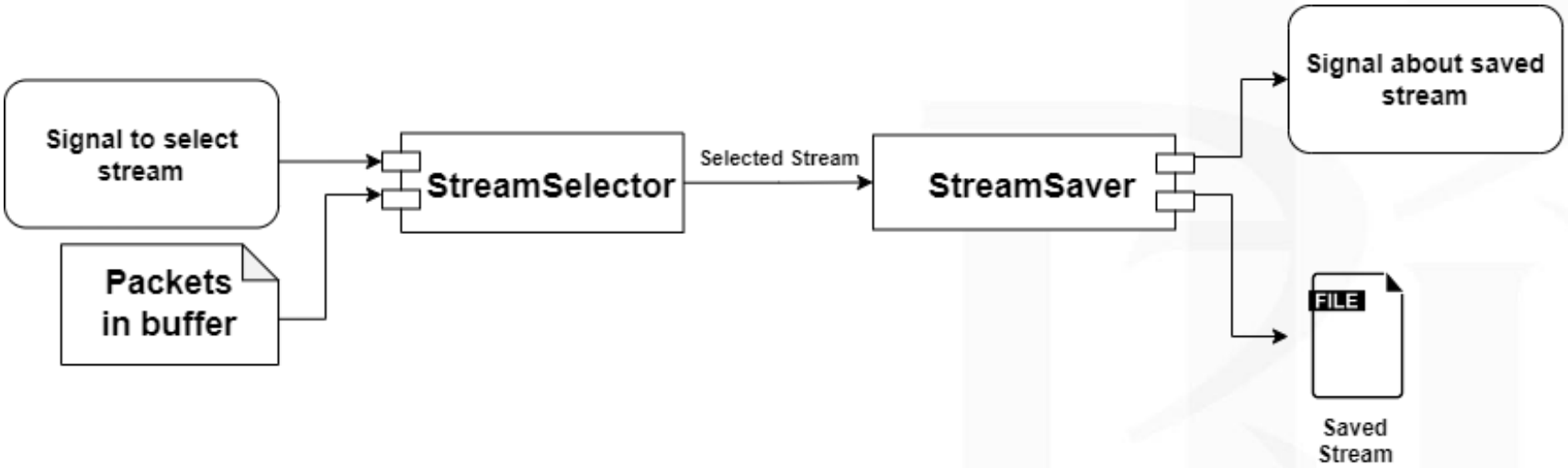
ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ



Из отфильтрованных на предыдущем шаге пакетов извлекается информация. Вначале проверяется тип установления соединения с помощью представления первых 4х символов из полезной нагрузки пакета в виде числа (одному символу соответствует один байт) и сравнения с возможными значениями.

После определения типа выполняется перемещение указателя внутри полезной нагрузки на начало информации о переданном адресе. Считанный адрес упаковывается в сигналы и отсылается различным модулям.

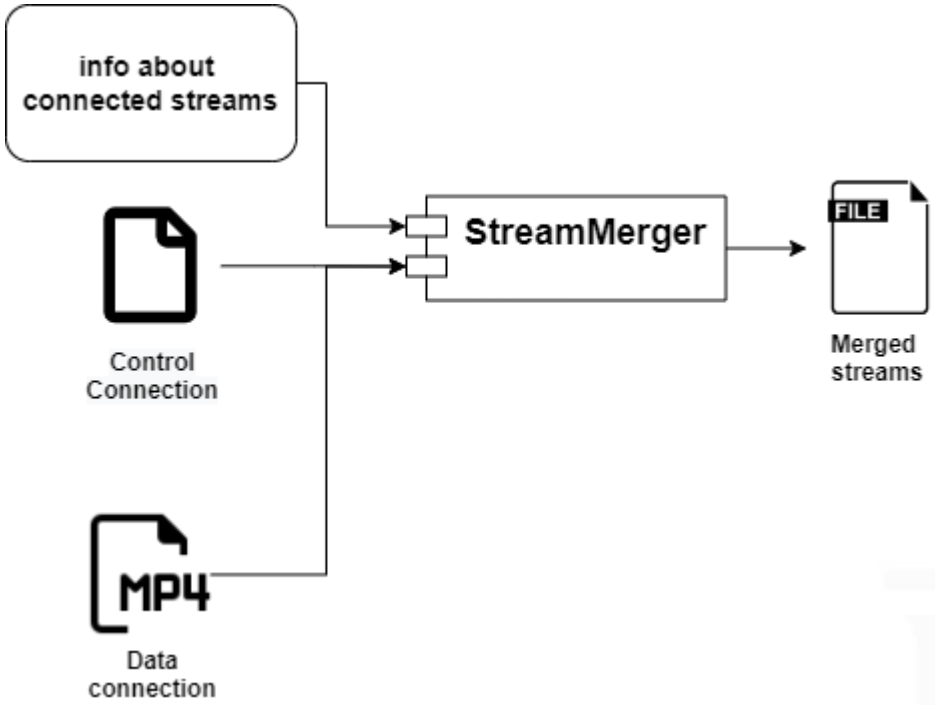
СОХРАНЕНИЕ ПАКЕТОВ ПОТОКА



При нахождении полной адресной информации, отправляется сигнал на отбор пакетов потока передаваемые другому модулю, который начинает сохранять их в pcap файл и дожидается пакета с меткой об окончании потока. Когда такой пакет приходит, то модуль высвобождает используемые ресурсы и отправляет сигнал об успешном сохранении.

Пример имени файла:
tcp_192.168.50.102_61692_192.168.50.75_21_16489769913624160000

ОБЪЕДИНЕНИЕ ПОТОКОВ



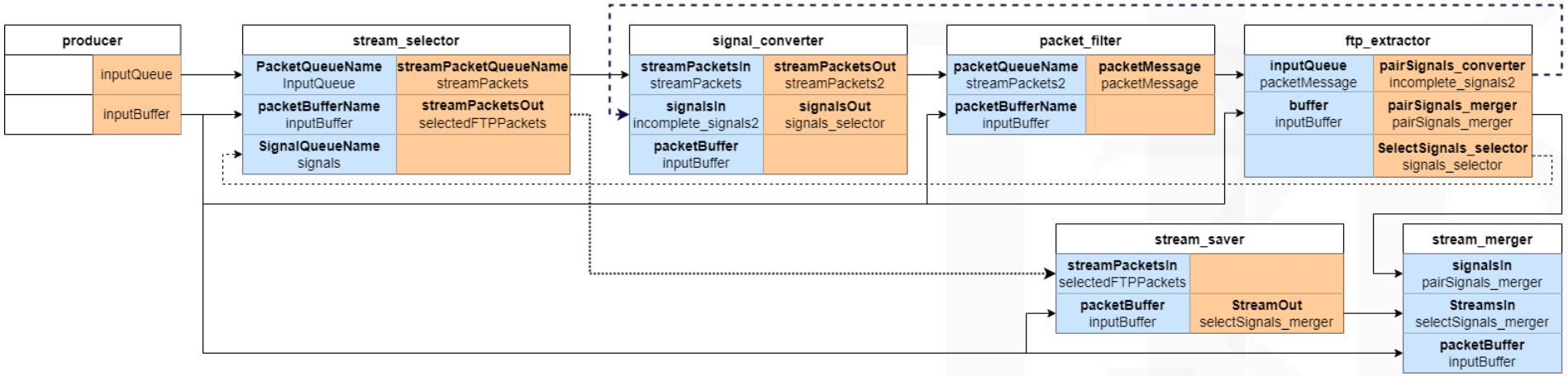
Информация о связанности потоков приходит в виде парного сигнала, который содержит информацию о потоке соединения данных и его родительском потоке управляющего соединения.

Данные этих сигналов сохраняются для последующего объединения

Само объединение может начаться только после получение сигнала о сохранении pcap файла. При получении такового, проверяется дошли ли все сигналы соединений данных и управляющего соединения, и если да, то начнётся объединение.

ДЕМОНСТРАЦИЯ

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ



В результате выполнения работы удалось собрать систему для анализа трафика использующего протокол передачи данных FTP. Данная система позволяет выделять и сохранять активность клиентов в pcap файлы. Каждому pcap файлу соответствует одно управляющее соединение и порождаемые им соединения данных.

ВОЗМОЖНОСТИ ДАЛЬНЕЙШЕГО РАЗВИТИЯ

Добавить возможность строить и использовать более сложные структуры иерархии.

Добавить возможность передачи объединенных потоков другим модулям с помощью записи в выходную очередь или отправки сигналов модулю, который считывает пакеты.

Разработанные модули и настроенное взаимодействие между ними могут быть включены в состав уже существующих или будущих решений для анализа сетевого трафика, тем самым расширив возможный функционал.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1) Джеймс Куроуз. Компьютерные сети. Нисходящий подход. // Джеймс Куроуз, Кит Росс – 6-е изд., Москва, 2016
2) Э. Таненбаум. Компьютерные Сети // Э. Таненбаум, Д. Уэзеролл – 5-ое изд., СПб.: Питер, 2012
3) Роберт Лав. Linux. Системное программирование // 2-е изд., Москва, 2018
4) Документация языка программирования C++. [Электронный ресурс]// URL: https://en.cppreference.com/
5) Hun-Jeong Kang, Myung-Sup Kim, and James Won-Ki Hong, “A Method on Multimedia Service Traffic Monitoring and Analysis”, Oct 2003

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Спасибо за внимание!

Глущенко Захар Сергеевич
zsgluschenko@edu.hse.ru
zsgluschenko@gmail.com

Москва - 2022