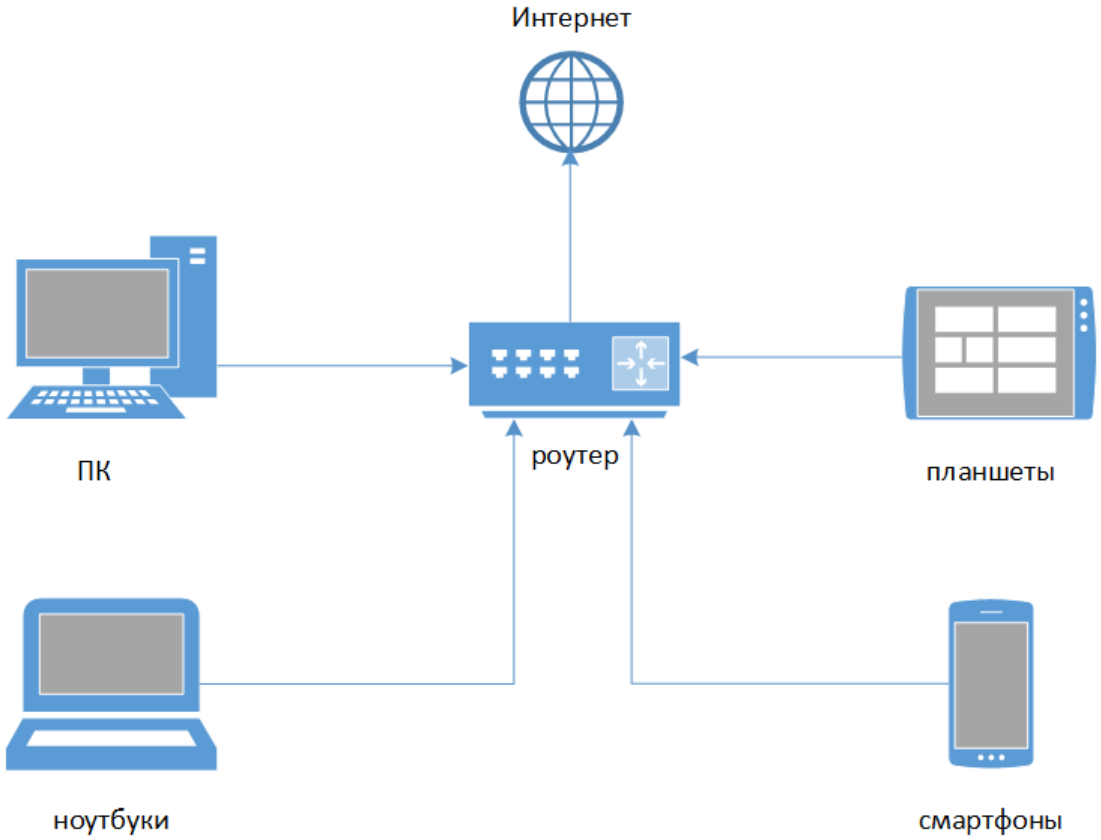
ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

ИВАНОВ ВАЛЕРИЙ В.
ТАТАРИНОВ И.И.

ТИПЫ КОМПЬЮТЕРНЫХ СЕТЕЙ: PAN

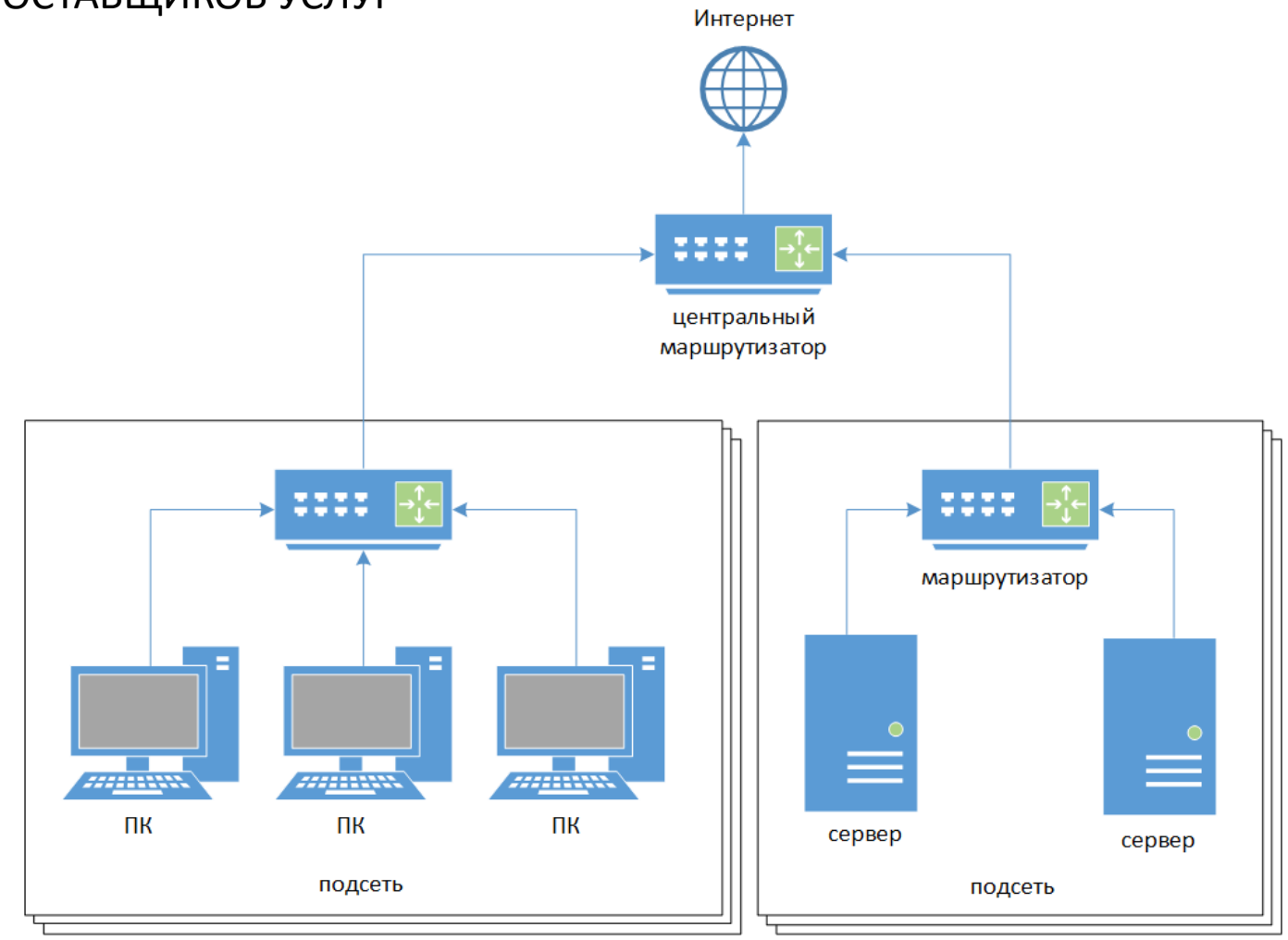
PAN (PERSONAL AREA NETWORK) — ПЕРСОНАЛЬНАЯ СЕТЬ, ПРЕДНАЗНАЧЕННАЯ ДЛЯ ВЗАИМОДЕЙСТВИЯ РАЗЛИЧНЫХ УСТРОЙСТВ, ПРИНАДЛЕЖАЩИХ ОДНОМУ ВЛАДЕЛЬЦУ



ТИПЫ КОМПЬЮТЕРНЫХ СЕТЕЙ: LAN

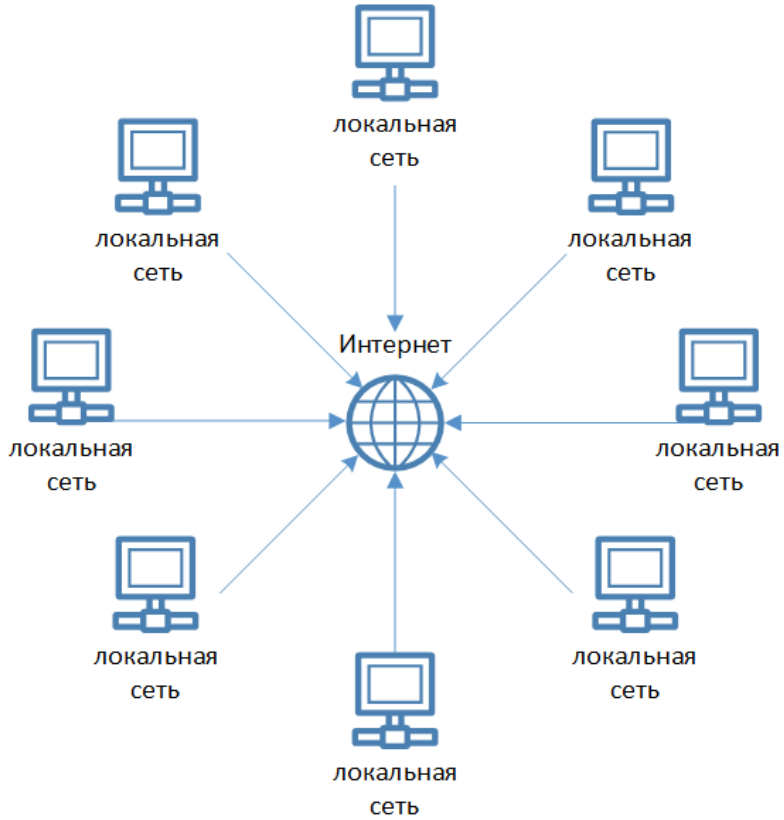
ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

LAN (ЛВС, LOCAL AREA NETWORK) — ЛОКАЛЬНАЯ СЕТЬ, ИМЕЮЩАЯ ЗАМКНУТУЮ ИНФРАСТРУКТУРУ ДО ВЫХОДА НА ПОСТАВЩИКОВ УСЛУГ



ТИПЫ КОМПЬЮТЕРНЫХ СЕТЕЙ: WAN

WAN (WIDE AREA NETWORK) — ГЛОБАЛЬНАЯ СЕТЬ, ВКЛЮЧАЮЩАЯ В СЕБЯ КАК ЛОКАЛЬНЫЕ СЕТИ, ТАК И ПРОЧИЕ ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ И УСТРОЙСТВА.



СЕТЕВЫЕ ПАКЕТЫ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

ОСНОВНЫЕ

РАЗМЕР

ВРЕМЯ ПЕРЕХВАТА

ВСПОМОГАТЕЛЬНЫЕ

ВХОДНОЙ MAC-АДРЕС УЗЛА ПЕРЕХВАТА

ВЫХОДНОЙ MAC-АДРЕС УЗЛА ПЕРЕХВАТА

IP-АДРЕС ОТПРАВИТЕЛЯ

IP-АДРЕС ПОЛУЧАТЕЛЯ

ТИП

ВТОРОСТЕПЕННЫЕ

ВРЕМЯ ЖИЗНИ (TTL)

ИДЕНТИФИКАТОР ПАКЕТА

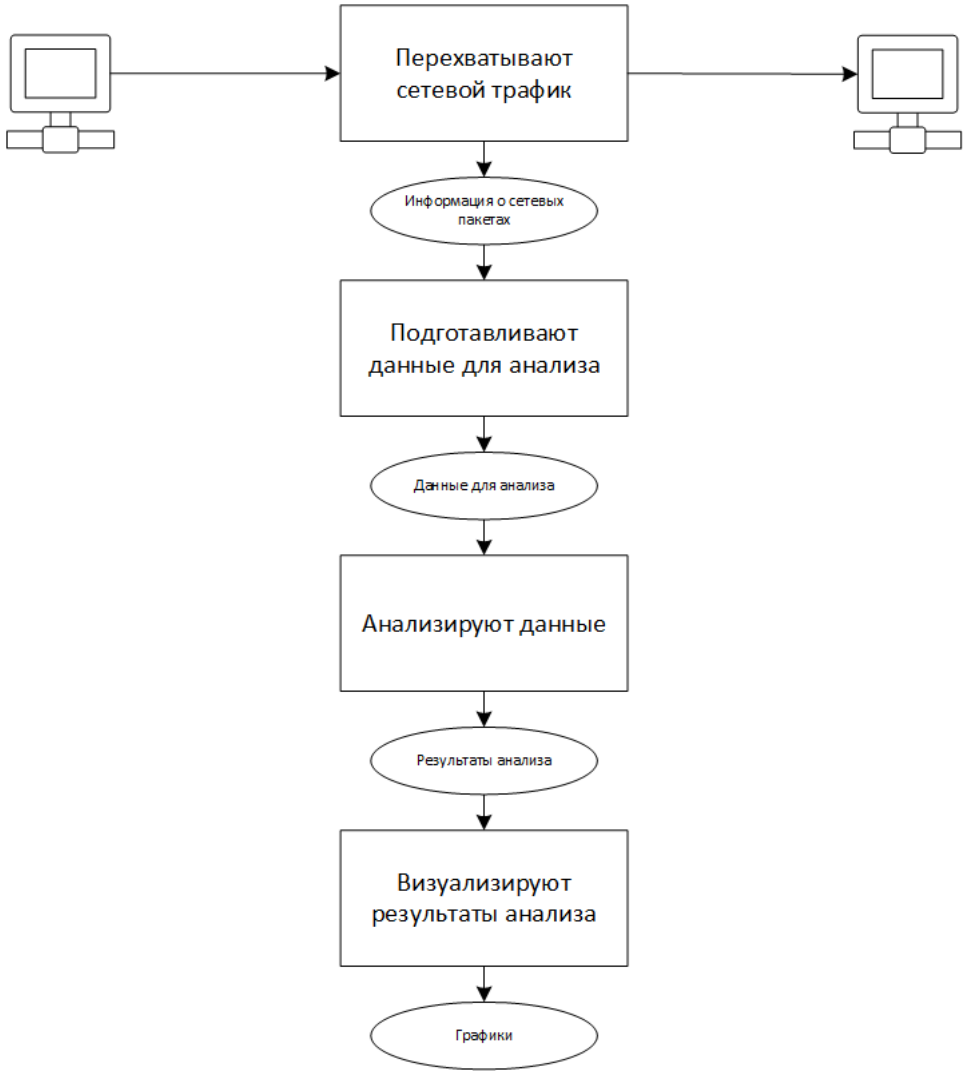
КОНТРОЛЬНАЯ СУММА

ФЛАГИ

ПРОЧИЕ СЛУЖЕБНЫЕ ДАННЫЕ

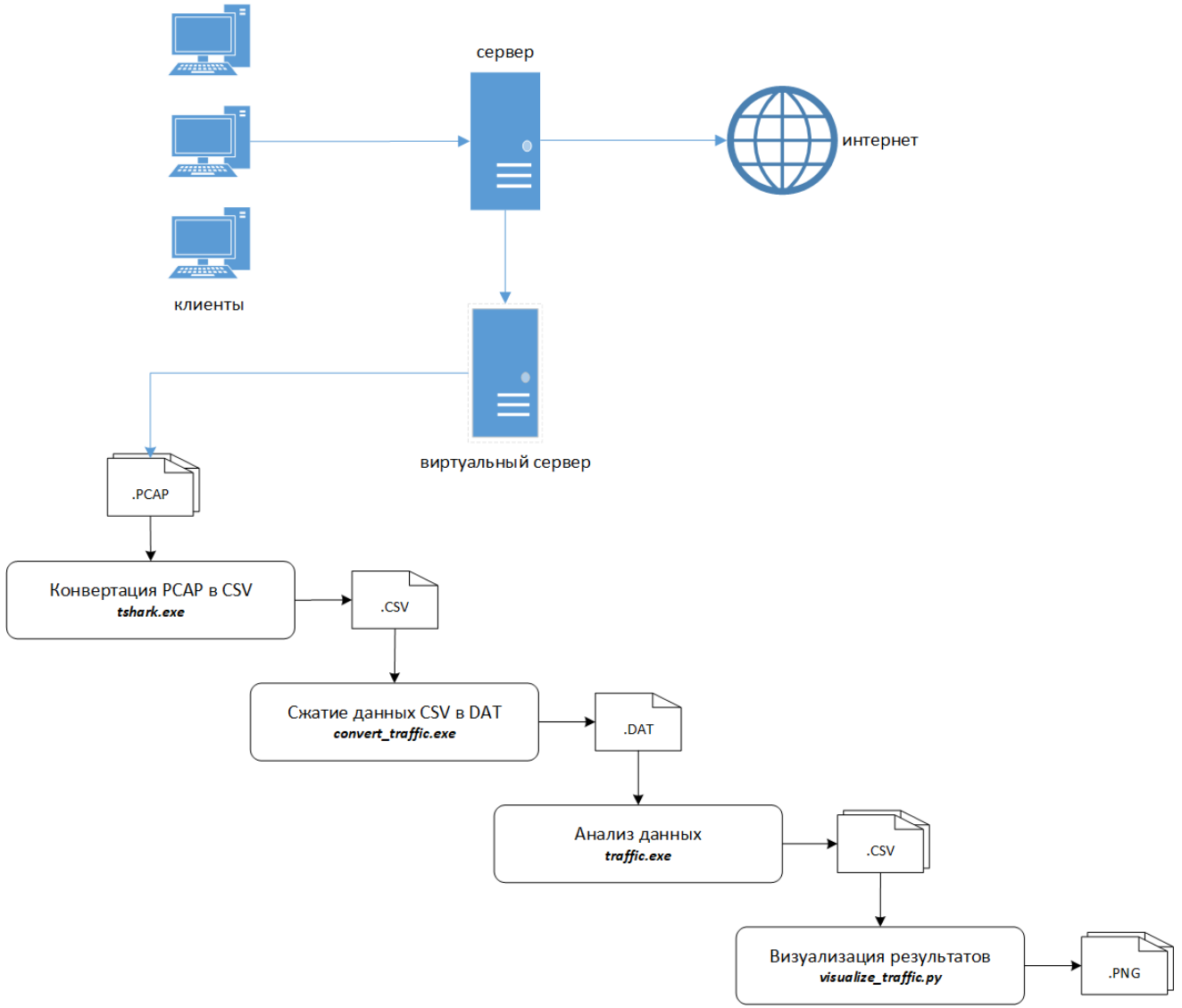
ИНСТРУМЕНТАРИЙ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА



ИНСТРУМЕНТАРИЙ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА



ПЕРЕХВАТ СЕТЕВОГО ТРАФИКА: ПО

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

ПРЕДОБРАБОТКА СЕТЕВЫХ ПАКЕТОВ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

НЕДОСТАТКИ ХРАНЕНИЯ ДАННЫХ О СЕТЕВЫХ ПАКЕТАХ В ROW-ФОРМАТЕ (.PCAP)

- ИЗБЫТОЧНЫЙ РАЗМЕР СЕТЕВОГО ПАКЕТА
- ПЛАВАЮЩИЙ (ПЕРЕМЕННЫЙ) РАЗМЕР СЕТЕВОГО ПАКЕТА
- АЛГОРИТМИЧЕСКАЯ СЛОЖНОСТЬ В ИЗВЛЕЧЕНИИ НЕОБХОДИМЫХ ДЛЯ АНАЛИЗА ДАННЫХ

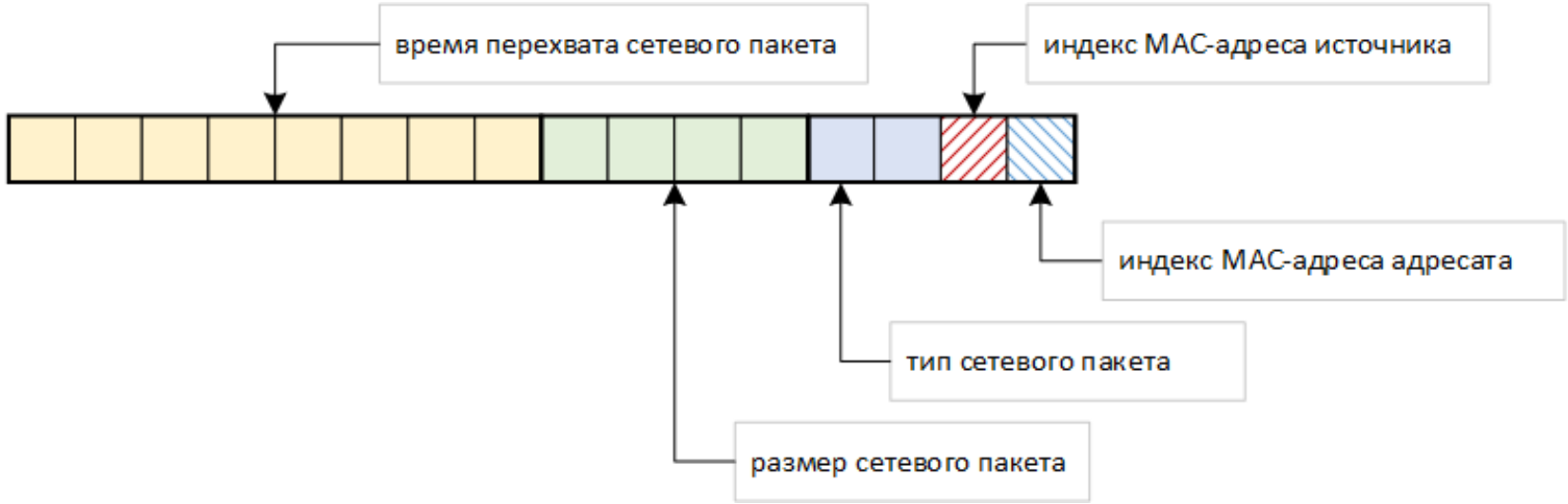
РЕШЕНИЕ

- ХРАНЕНИЕ ЛИШЬ НЕОБХОДИМЫХ ДЛЯ АНАЛИЗА ДАННЫХ
- КОМПАТИФИКАЦИЯ ДАННЫХ (ИСПОЛЬЗОВАНИЕ БИНАРНОГО ПРЕДСТАВЛЕНИЯ И ИНДЕКСОВ)
- ФИКСИРОВАННЫЙ РАЗМЕР ДАННЫХ
- БЛОЧНАЯ АРХИВАЦИЯ ДАННЫХ

ПРЕДОБРАБОТКА СЕТЕВЫХ ПАКЕТОВ

ТИП ДАННЫХ 1

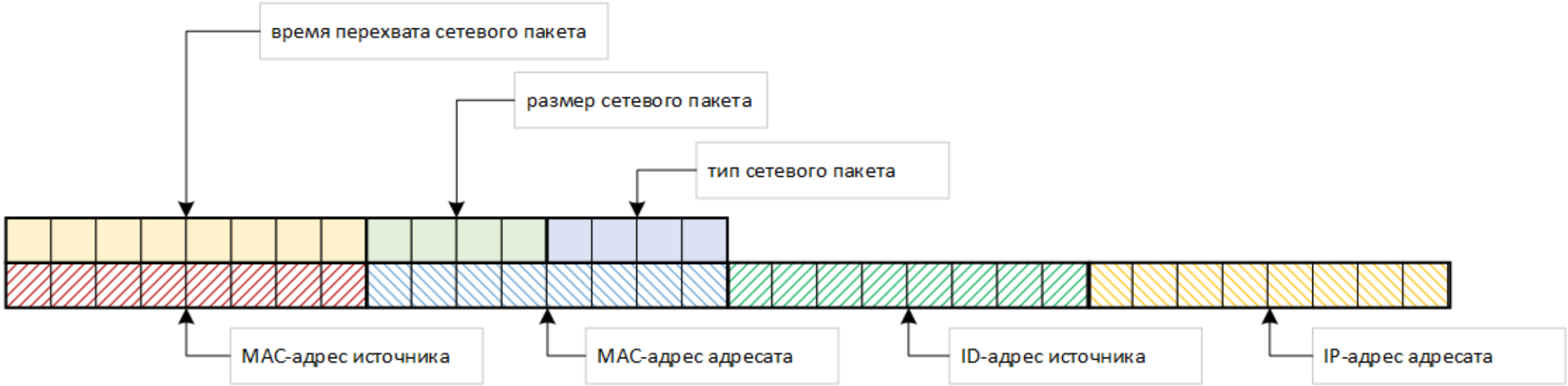
16 БАЙТ НА 1 СЕТЕВОЙ ПАКЕТ



ПРЕДОБРАБОТКА СЕТЕВЫХ ПАКЕТОВ

ТИП ДАННЫХ 2

48 БАЙТ НА 1 СЕТЕВОЙ ПАКЕТ



ПРЕДОБРАБОТКА: СТАТИСТИКА

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

ДЛИТЕЛЬНОСТЬ СЕТЕВОГО ТРАФИКА	48 ЧАСОВ
КОЛИЧЕСТВО СЕТЕВЫХ ПАКЕТОВ	15+ МИЛЛИАРДОВ
ОБЪЁМ «СЫРЫХ» ДАННЫХ	1172 ГБ
ОБЪЁМ ПРЕДОБРАБОТАННЫХ ДАННЫХ	228 ГБ
ОБЪЁМ СЖАТЫХ ДАННЫХ	72ГБ

ОСНОВНОЙ ИНСТРУМЕНТАРИЙ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

НАКОПЛЕНИЕ СТАТИСТИКИ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

ХАРАКТЕРИСТИКИ СЕТЕВЫХ ПАКЕТОВ

- РАЗМЕРЫ СЕТЕВЫХ ПАКЕТОВ
- ВРЕМЯ МЕЖДУ СОСЕДНИМИ СЕТЕВЫМИ ПАКЕТАМИ

ВРЕМЕННЫЕ ХАРАКТЕРИСТИКИ СЕТЕВОГО ТРАФИКА

- ИНТЕНСИВНОСТЬ СЕТЕВОГО ТРАФИКА ВО ВРЕМЕНИ
- СКОРОСТЬ ПЕРЕДАЧИ СЕТЕВЫХ ПАКЕТОВ ВО ВРЕМЕНИ

ПРЕДЕЛЬНЫЕ ВЕРОЯТНОСТНЫЕ РАСПРЕДЕЛЕНИЯ

- ИНТЕНСИВНОСТЬ СЕТЕВОГО ТРАФИКА
- СКОРОСТЬ ПЕРЕДАЧИ СЕТЕВЫХ ПАКЕТОВ

НАКОПЛЕНИЕ СТАТИСТИКИ: ПРОБЛЕМАТИКА

ПРОБЛЕМЫ НАКОПЛЕНИЯ СТАТИСТИКИ

ОГРОМНЫЕ ОБЪЁМЫ ДАННЫХ ДЛЯ АНАЛИЗА

НЕВОЗМОЖНО ОБРАБОТАТЬ ВСЕ ДАННЫЕ ЗА ОДИН ПРОХОД (НЕ ХВАТАЕТ ОПЕРАТИВНОЙ ПАМЯТИ)

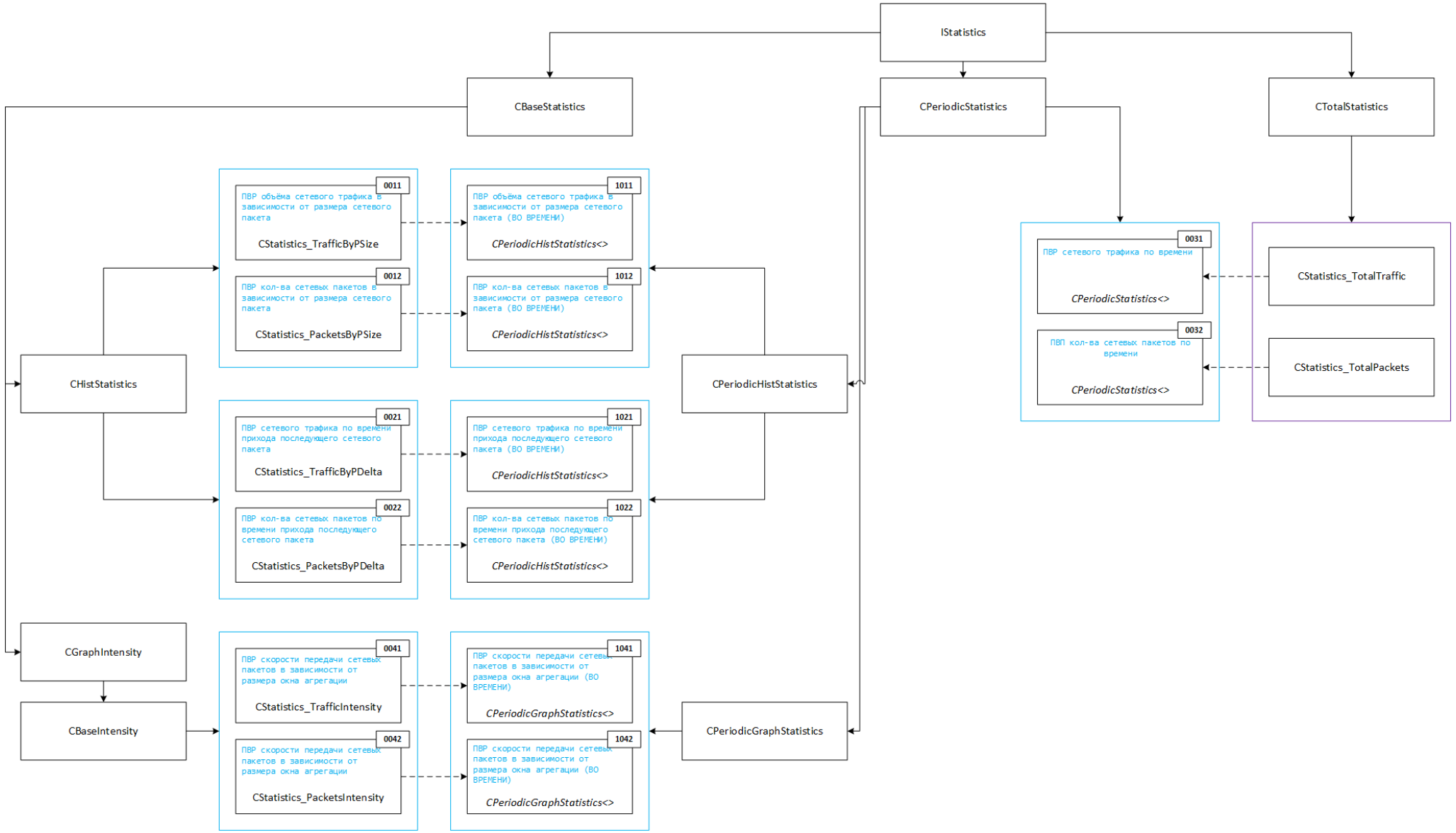
РЕШЕНИЕ

ИСПОЛЬЗОВАНИЕ КОМПАТИФИЦИРОВАННЫХ ДАННЫХ

СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ ДАННЫХ ДЛЯ РАЗНЫХ РАСПРЕДЕЛЕНИЙ

НАКОПЛЕНИЕ СТАТИСТИКИ: ПО

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

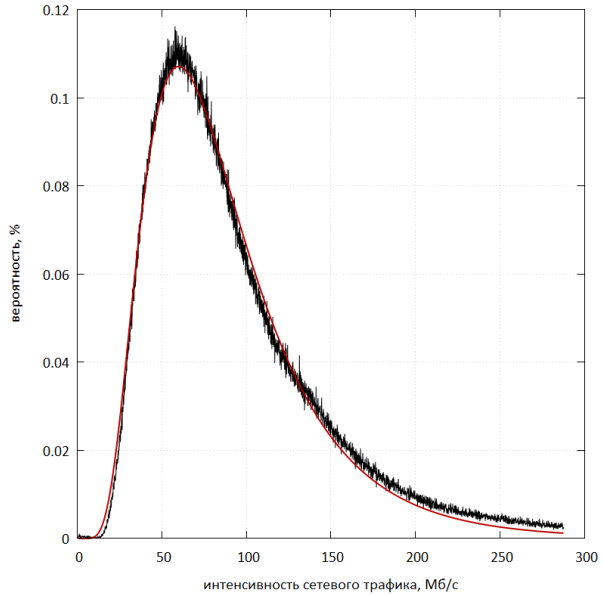


АППРОКСИМАЦИЯ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

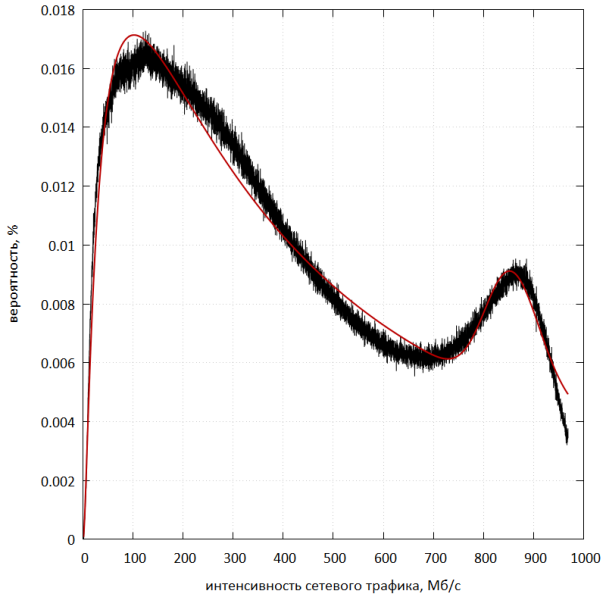
АППРОКСИМАЦИЯ ЛОГНОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ

$$f(t) = \frac{a}{\sqrt{2\pi}\sigma t} e^{\frac{(\ln t - \mu)^2}{2\sigma^2}}$$



АППРОКСИМАЦИЯ СУММОЙ НЕСКОЛЬКИХ ЛОГНОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ

$$f(t) = \sum \frac{a_i}{\sqrt{2\pi}\sigma_i t} e^{\frac{(\ln t - \mu_i)^2}{2{\sigma_i}^2}}$$



АППРОКСИМАЦИЯ: АЛГОРИТМ

ВХОДНЫЕ ДАННЫЕ $(t_1, p_1), (t_2, p_2), \dots, (t_N, p_N)$

РЕЗУЛЬТАТ АППРОКСИМАЦИИ a, σ, μ

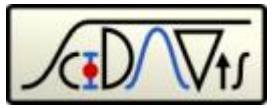
$$f(t) = \frac{a}{\sqrt{2\pi}\sigma t} e^{\frac{(\ln t - \mu)^2}{2\sigma^2}}$$

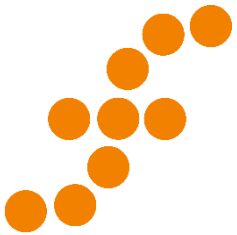
В КАЧЕСТВЕ КРИТЕРИЯ ТОЧНОСТИ АППРОКСИМАЦИИ ИСПОЛЬЗОВАЛСЯ
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

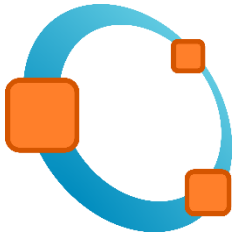
$$S_{ols} = \sum (p_i - f(t_i))^2 = min$$

АППРОКСИМАЦИЯ: ПО

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА







АППРОКСИМАЦИЯ: АЛГОРИТМ

$$f(t) = \frac{a}{\sqrt{2\pi}\sigma t} e^{\frac{(\ln t - \mu)^2}{2\sigma^2}}$$

σ, μ ВЫЧИСЛЯЛИСЬ МЕТОДАМИ ГРАДИЕНТНОГО СПУСКА

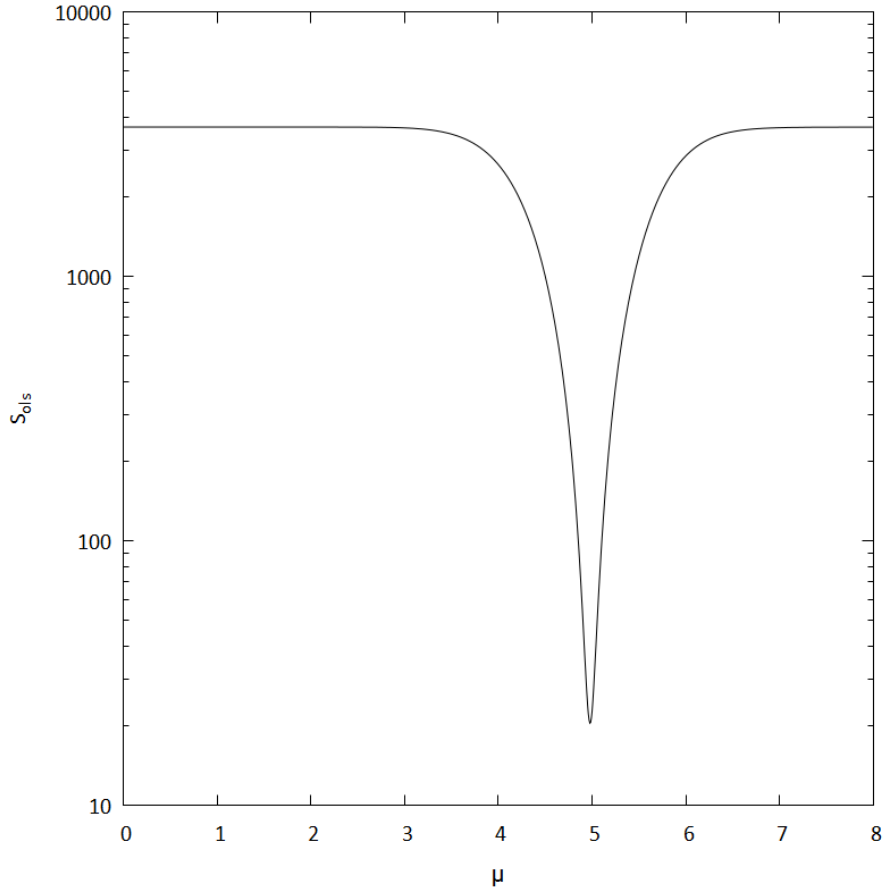
a ВЫЧИСЛЯЛСЯ АНАЛИТИЧЕСКИ КАК РЕШЕНИЕ УРАВНЕНИЯ

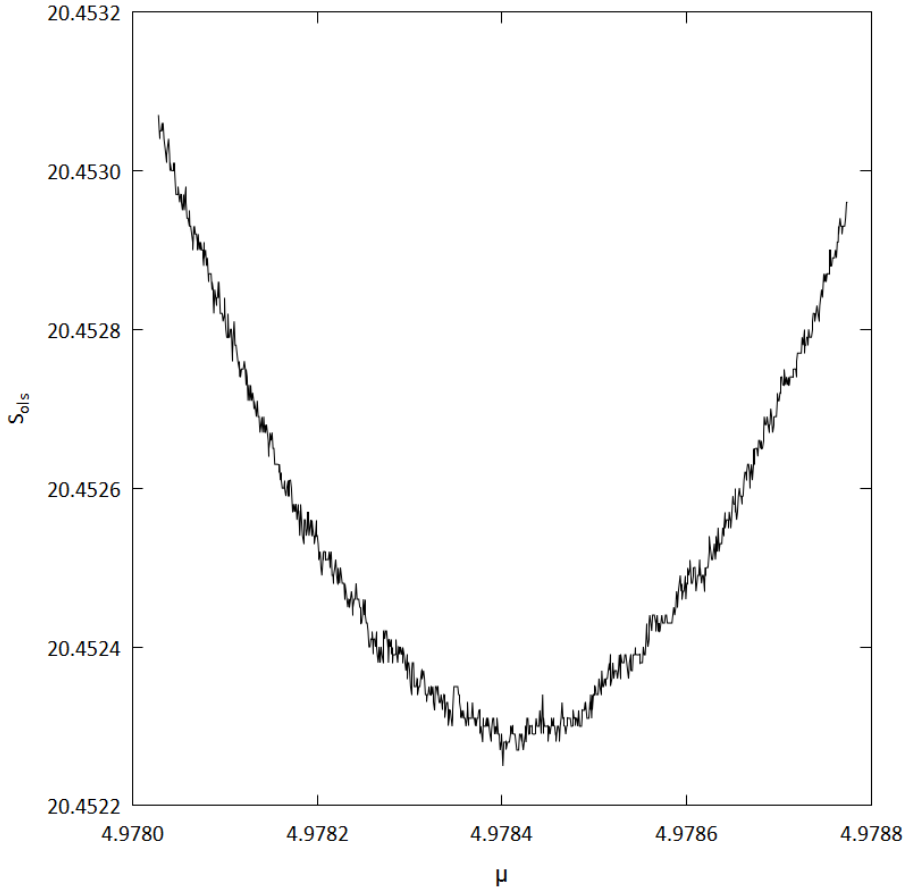
$$S_{ols} = \sum (p_i - f(t_i, \sigma, \mu))^2 = min$$

$$a = \frac{\sum p_i f(t_i, \sigma, \mu)}{\sum (f(t_i, \sigma, \mu))^2}$$

АППРОКСИМАЦИЯ: ГРАДИЕНТНЫЙ СПУСК

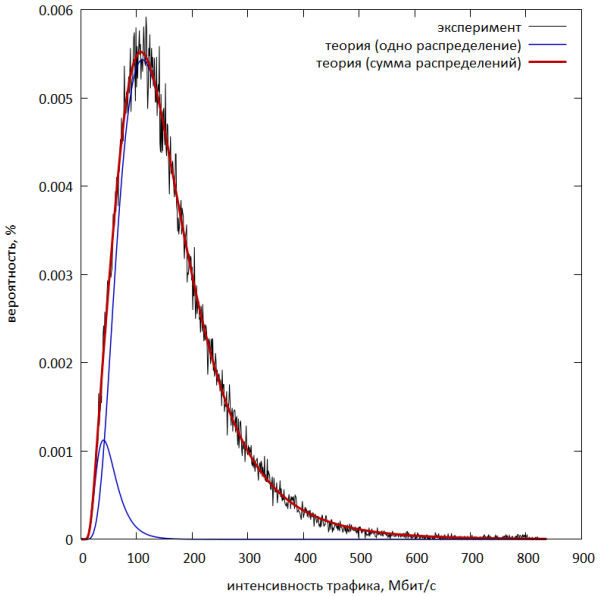
ПРОБЛЕМА ОВРАГОВ

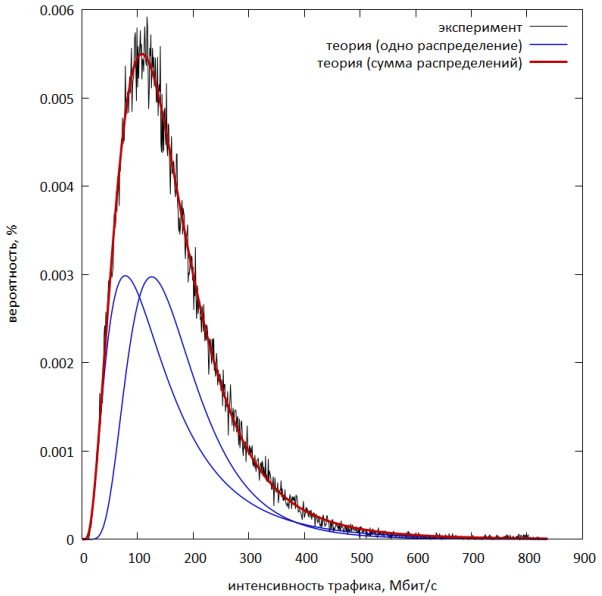




АППРОКСИМАЦИЯ: ГРАДИЕНТНЫЙ СПУСК

ПРОБЛЕМА: СХОЖИЙ ИТОГОВЫЙ РЕЗУЛЬТАТ ПРИ СИЛЬНО РАЗЛИЧНЫХ ХАРАКТЕРИСТИКАХ РАСПРЕДЕЛЕНИЙ

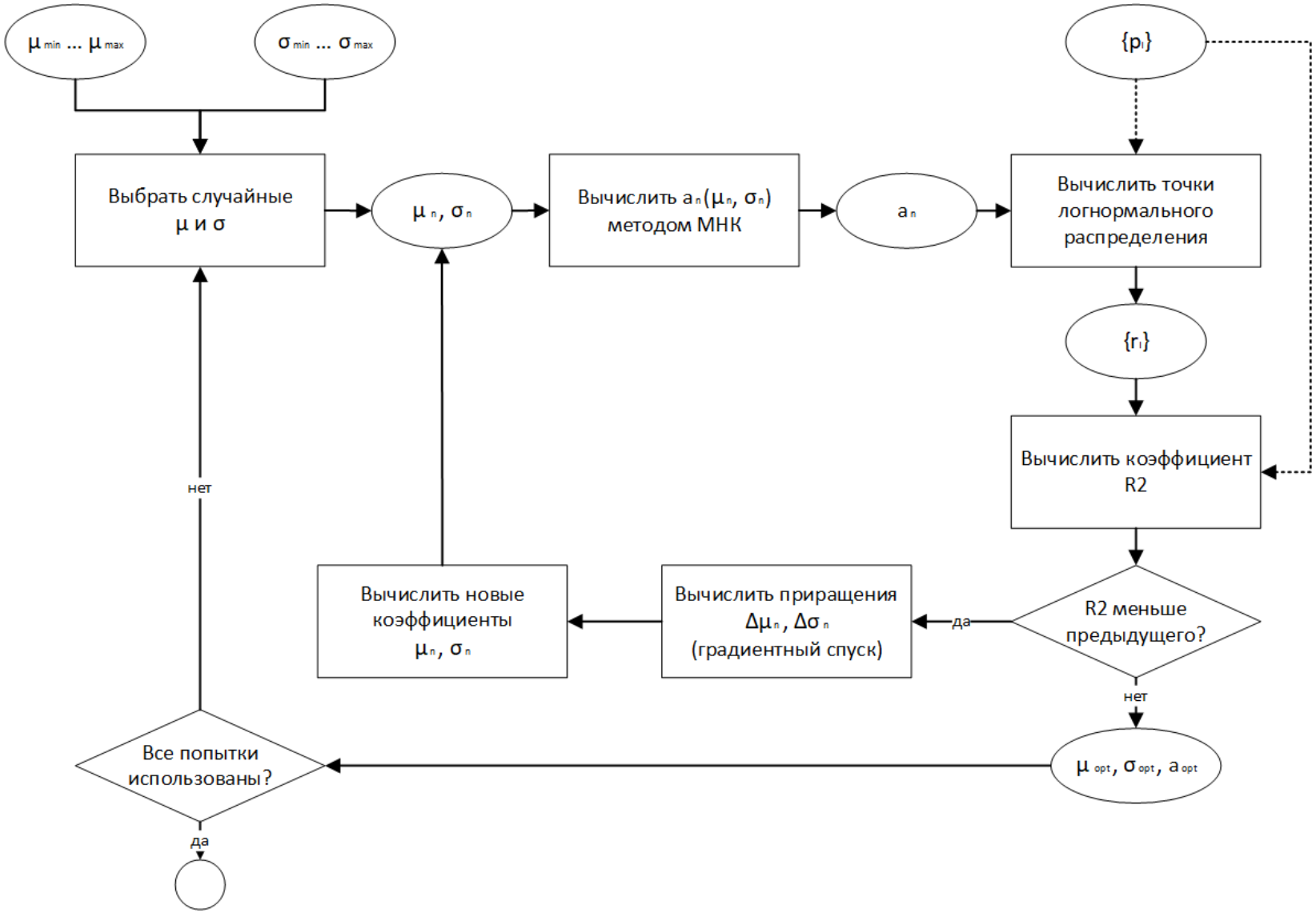




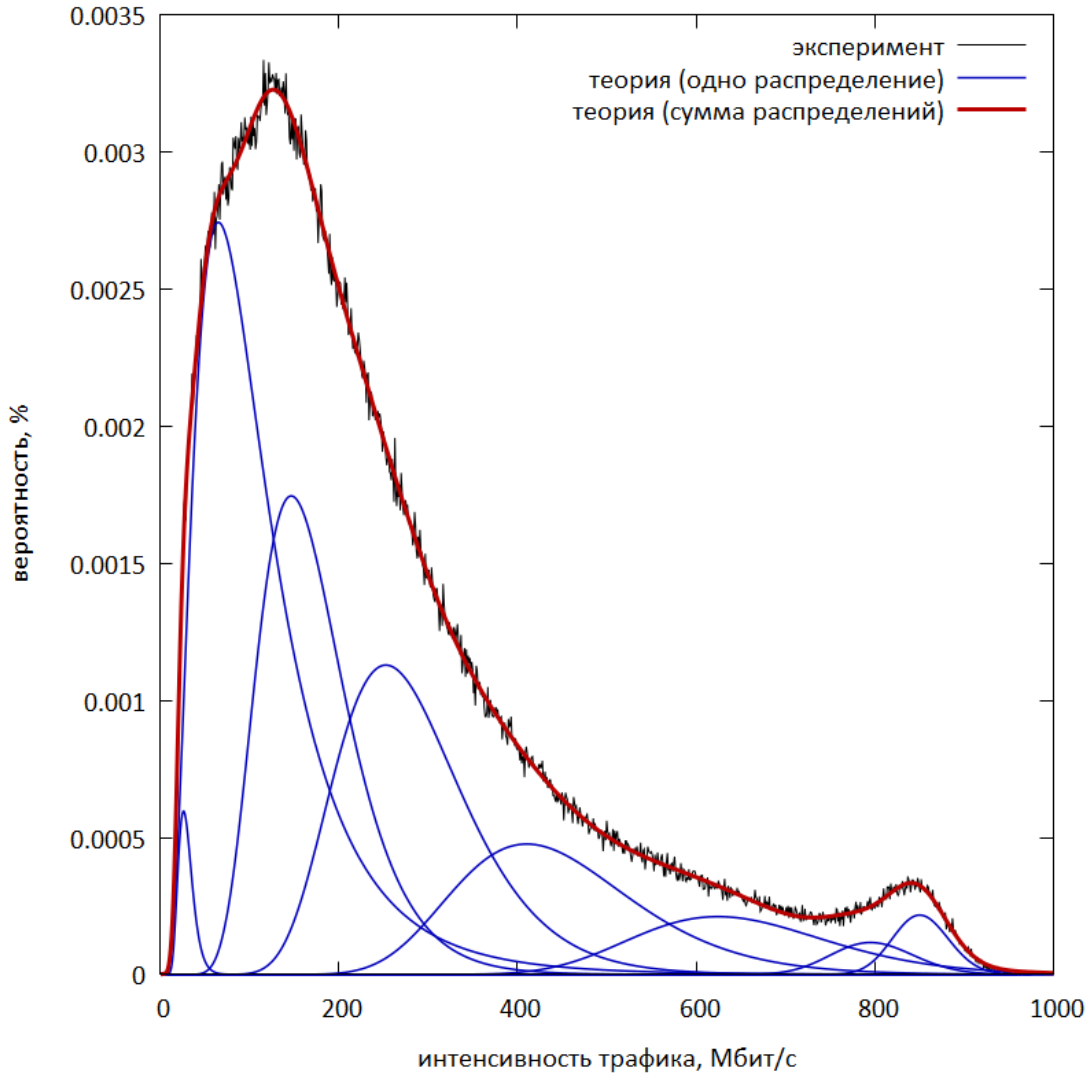
РЕШЕНИЕ: МНОЖЕСТВЕННЫЕ ПОПЫТКИ ВЫПОЛНИТЬ ГРАДИЕНТНЫЙ СПУСК С РАЗЛИЧНЫМИ НАЧАЛЬНЫМИ ПАРАМЕТРАМИ

АППРОКСИМАЦИЯ: ПО

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

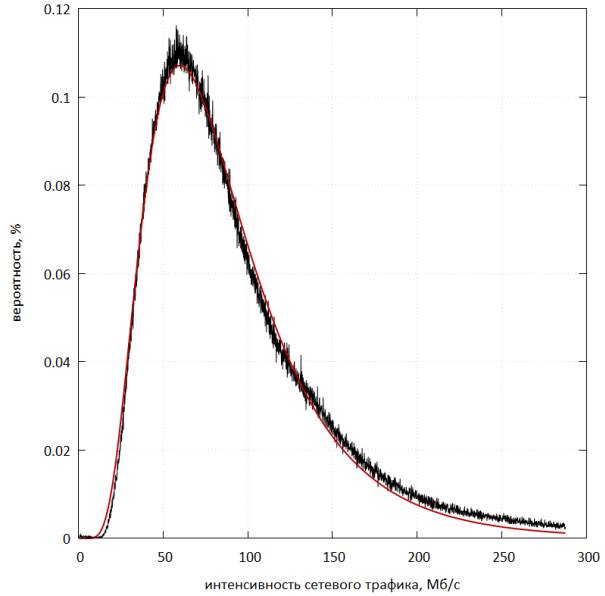


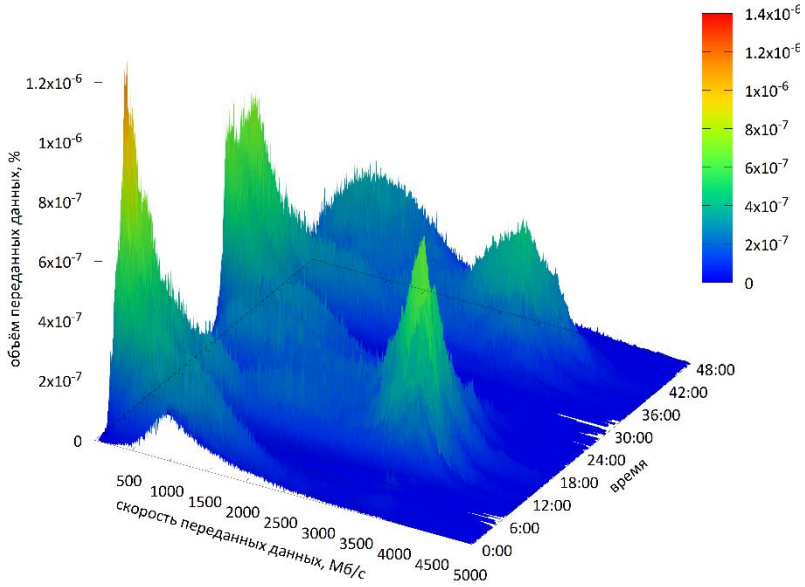
АППРОКСИМАЦИЯ: РЕЗУЛЬТАТЫ

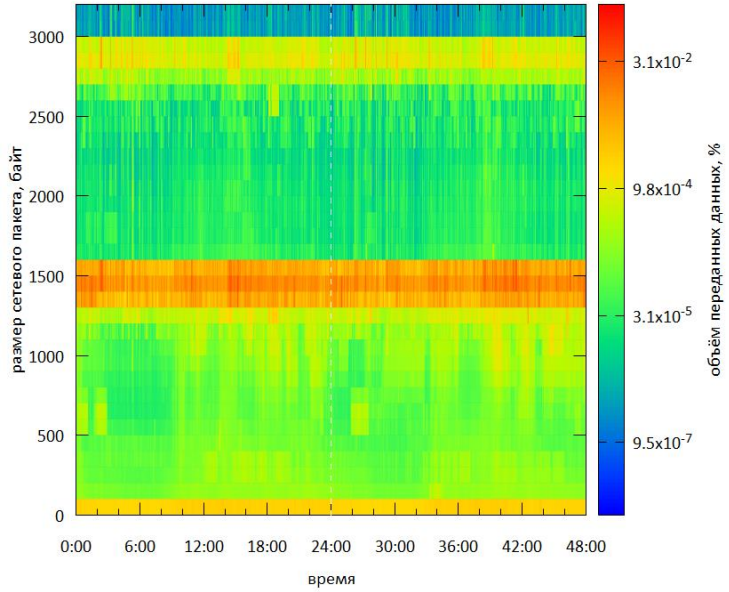


ВИЗУАЛИЗАЦИЯ

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА

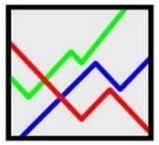






ВИЗУАЛИЗАЦИЯ: ПО

ОСОБЕННОСТИ ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА



gnuplot

Gnuplot
Scripts

python™

matplotlib

MESHMIXER

СПАСИБО

ВОПРОСЫ?