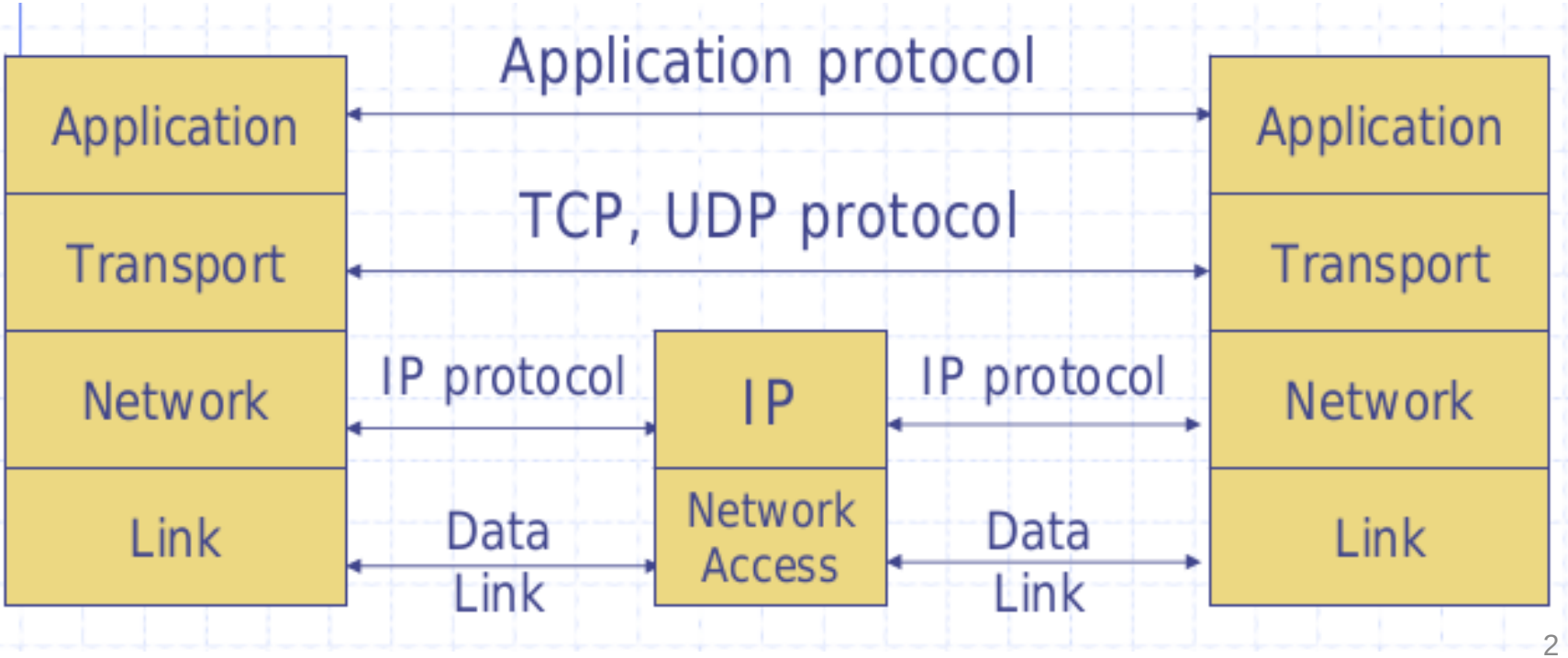
Анализ сетевого трафика

Гетьман Александр

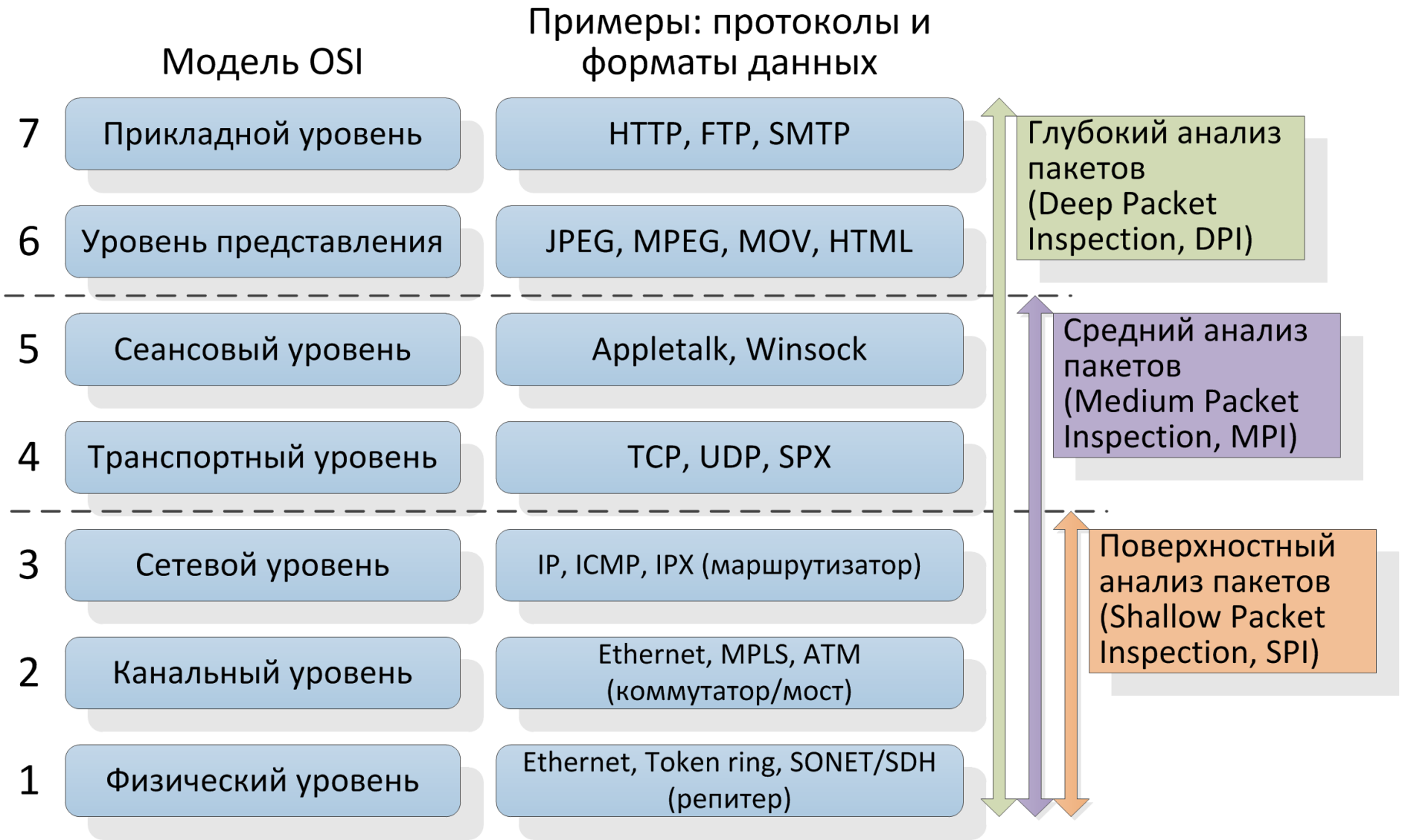
thorin@ispras.ru

Сетевой стек

- Модель OSI – 7 уровней. В процессе передаче используются нижние 3.
- Модель TCP/IP – 4 уровня. В процессе передаче используются нижние 2.
- Основа процесса передачи – маршрутизация, т.е. выбор сетевого порта, на который необходимо передать пакет в зависимости от значений его полей



Уровни детализации анализа трафика



Подзадачи DPI-анализа трафика

Прикладные задачи анализа трафика
- Анализ файлов (поиск вирусов), анализ текстов писем (спам)
- Анализ аудио/видео-потоков (оценка качества)

Разбор трафика как сервис

Разбор прикладных протоколов (L7)
- Распознавание протоколов
- Выделение данных уровня приложения

Управление обработкой потоков данных

Группировка, фильтрация, хранение (L3-L4)
- Группировка пакетов по потокам, связанные потоки
- Фильтрация по адресным признакам

Неблокирующая синхронизация

Получение данных из сетевого канала (L2)
- 10 Гбит/64 байта на пакет без потерь (1 порт)
- Параллельная обработка (SMP-машины)

Спец. аппаратура: сетевые карты, балансировщики

Постановки задачи анализа сетевого трафика

Особенности	Анализ на потоке	Анализ трасс
Прикладные задачи анализа	• Безопасность периметра (Firewall, IDS/IPS) • Качество связи, управление трафиком (QoS, Shaping) • Управление политиками (PCEF, NAC) • Безопасность внутри сети (Анализ поведения)	• Расследования инцидентов нарушения информационной безопасности • Анализ несоответствий между стандартом протокола и реализацией • Обратная инженерия и отладка сетевых протоколов

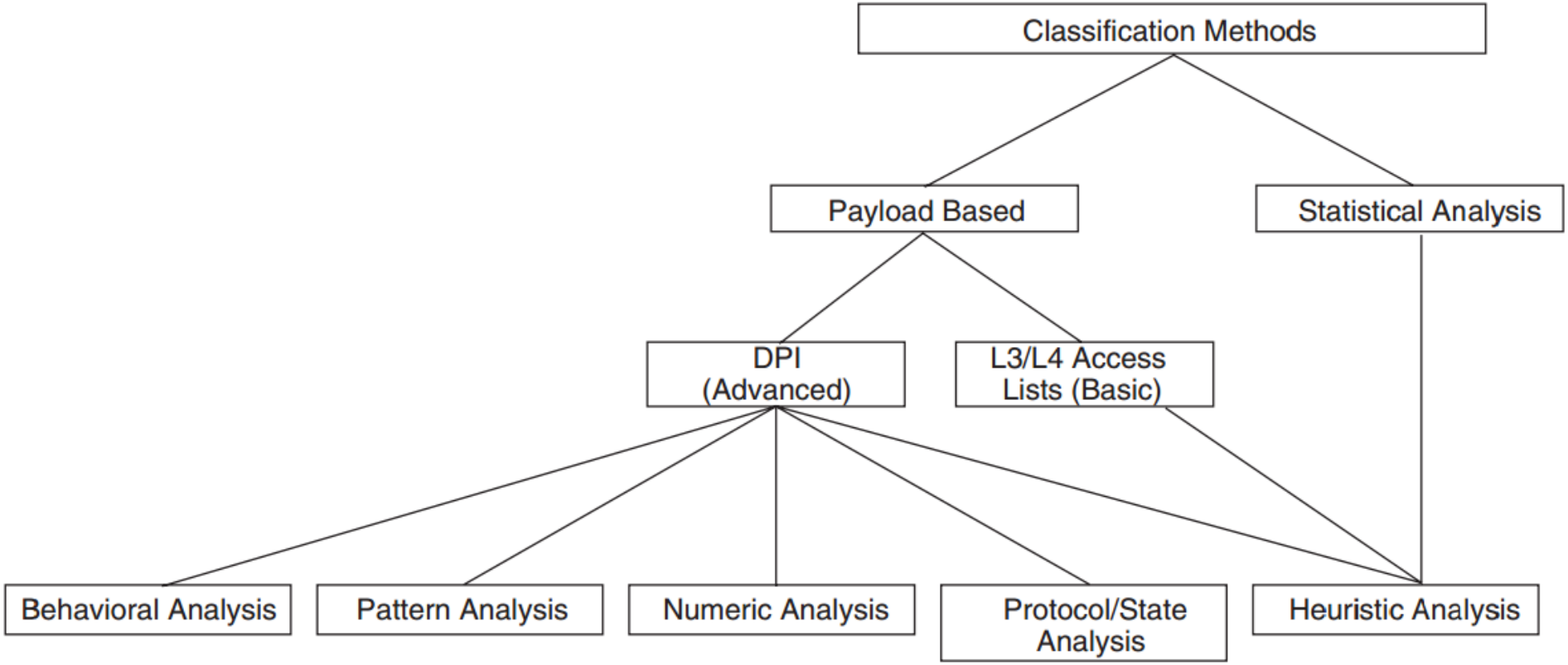
Условия анализа	• Ресурсы ограничены (память, процессор, пропускная способность шин)
	• Приоритет – скорость
		• Ресурсов достаточно
		  • Небольшие трассы
		  • Ленивая подгрузка
		• Приоритет – детализация

Доступные инструменты	• Большое количество • Каждый решает конкретную задачу	• Wireshark (Tshark) • Более 1500 протоколов

Перехват высокоскоростного трафика

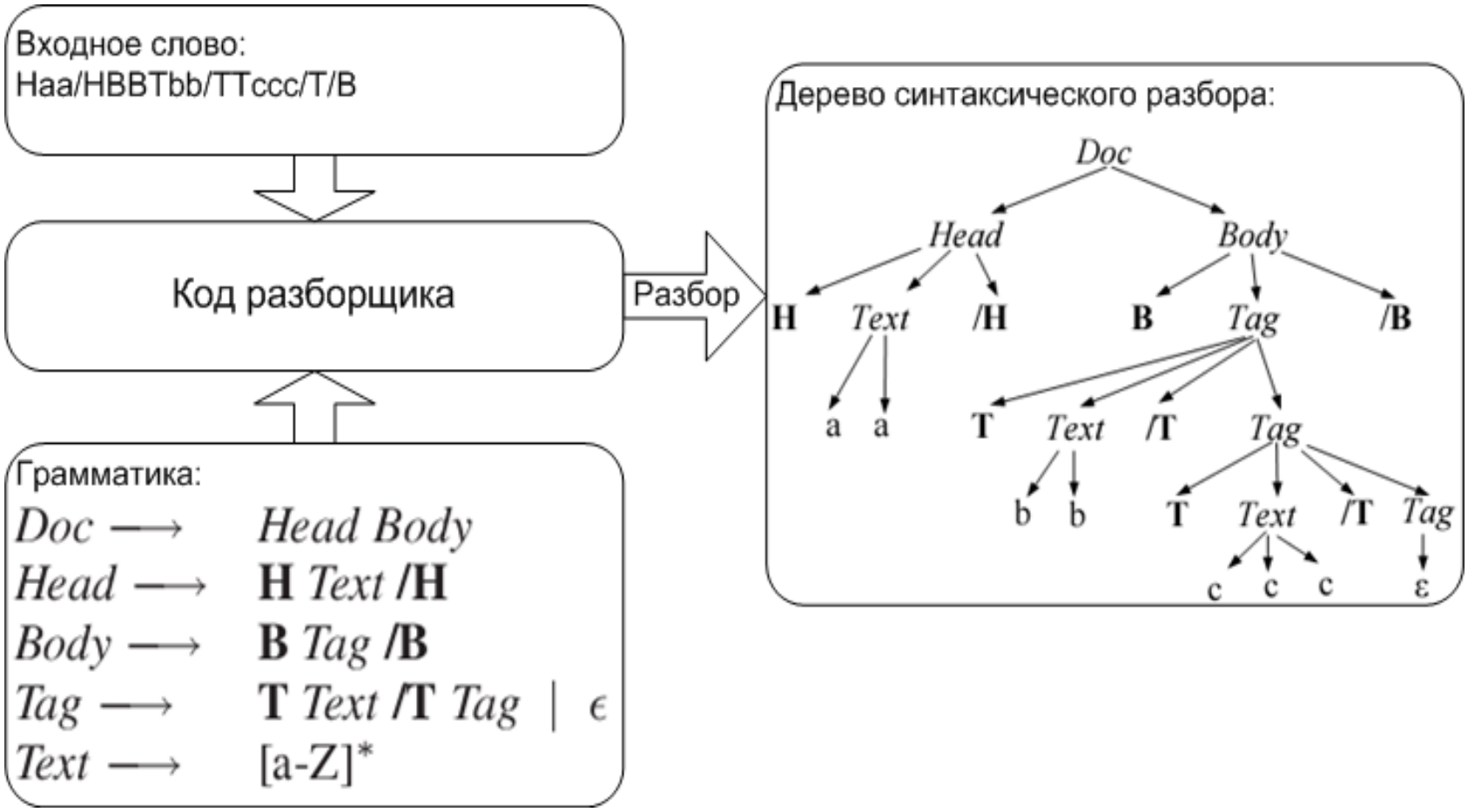
- Первый этап анализа потока на широком канале
- Сетевые каналы 10/40/100 GbE, 10 Гбит/сек = 1.25 Гбайт/сек = 14 880 952 пакетов/сек по 64 байта ~ 200 тактов на пакет
- Возникающие проблемы и подходы к решению:
  - Стандартные сетевые карты 10GbE со стандартными драйверами не позволяют перехватывать 100% пакетов. **Спец. карты и сетевые стеки**.
  - Большое количество копирований данных пакетов (сетевая карта -> буфер ядра -> буфер пользователя). **Подход zero-copy (DMA).**
  - Большое количество аппаратных прерываний, системных вызовов переключений контекстов. **Специализированные драйвера.**
  - Недостаточное использование параллелизма: все прерывания на 1 ядро. **Программно-аппаратные технологии, например MSI-X**
  - Проблемы синхронизации при параллельном доступе к данным сетевых пакетов из нескольких ядер. **Lock-free структуры данных.**

Основная задача всех DPI-систем: классификация трафика



- Неполный список применяемых методов:
  - Статистический анализ
  - Алгоритмы кластеризации
  - Конечные автоматы, регулярные выражения, грамматики
  - Машинное обучение
  - Методы автоматического извлечения сигнатур

Основная задача развитых DPI-систем: «извлечение метаинформации»

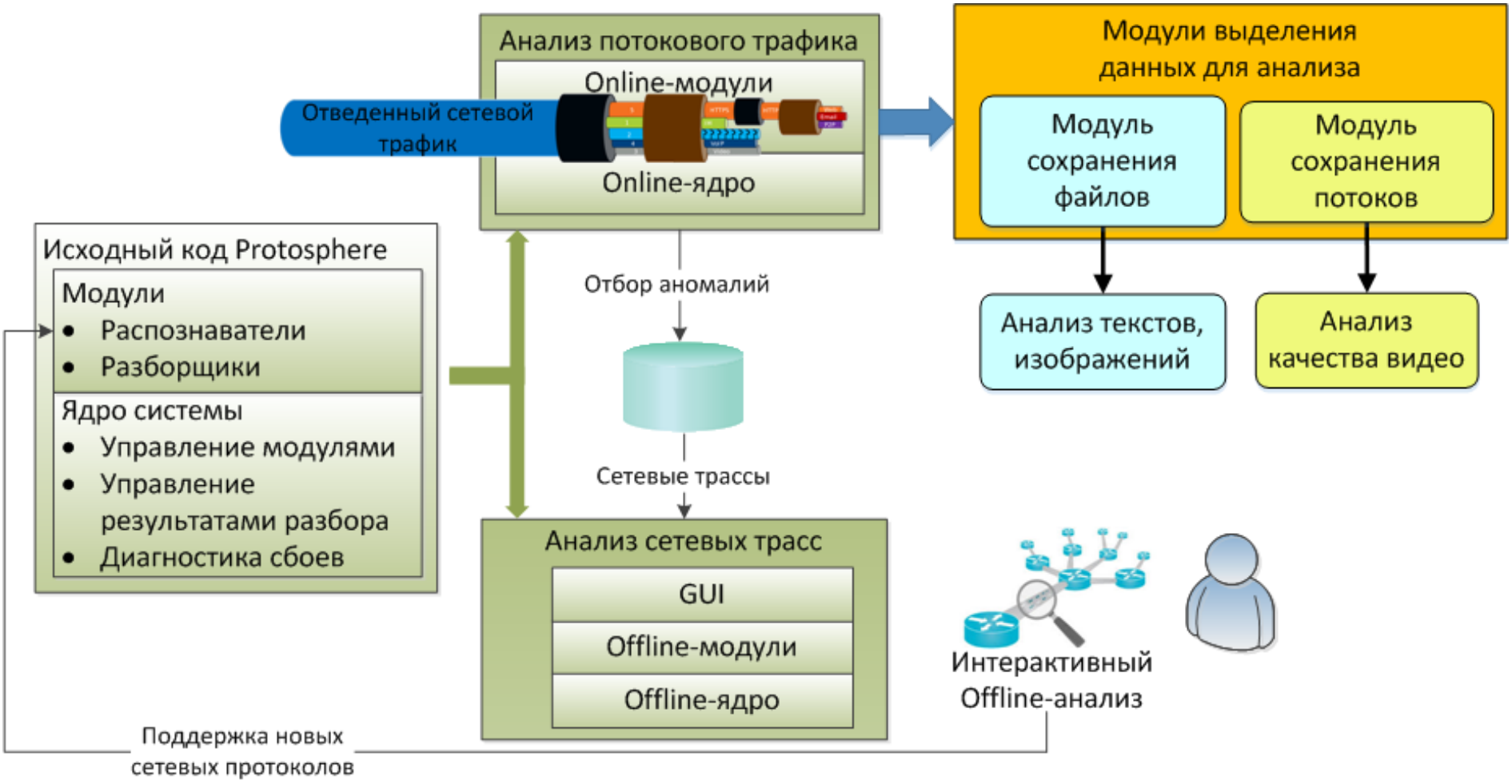


- Фактически – синтаксический разбор сообщений с извлечением значений полей. Методы
  - Теория языков. Декларативные и смешанные формы описания языков.
  - Специализированные типы деревьев, как форма представления результатов разбора
  - Теория компьютерных сетей: сетевые потоки, многопоточные соединения, автоматы протоколов, описание состояний сетевых сессий

Пример анализа шифрованного трафика: оценка характеристик потокового видео

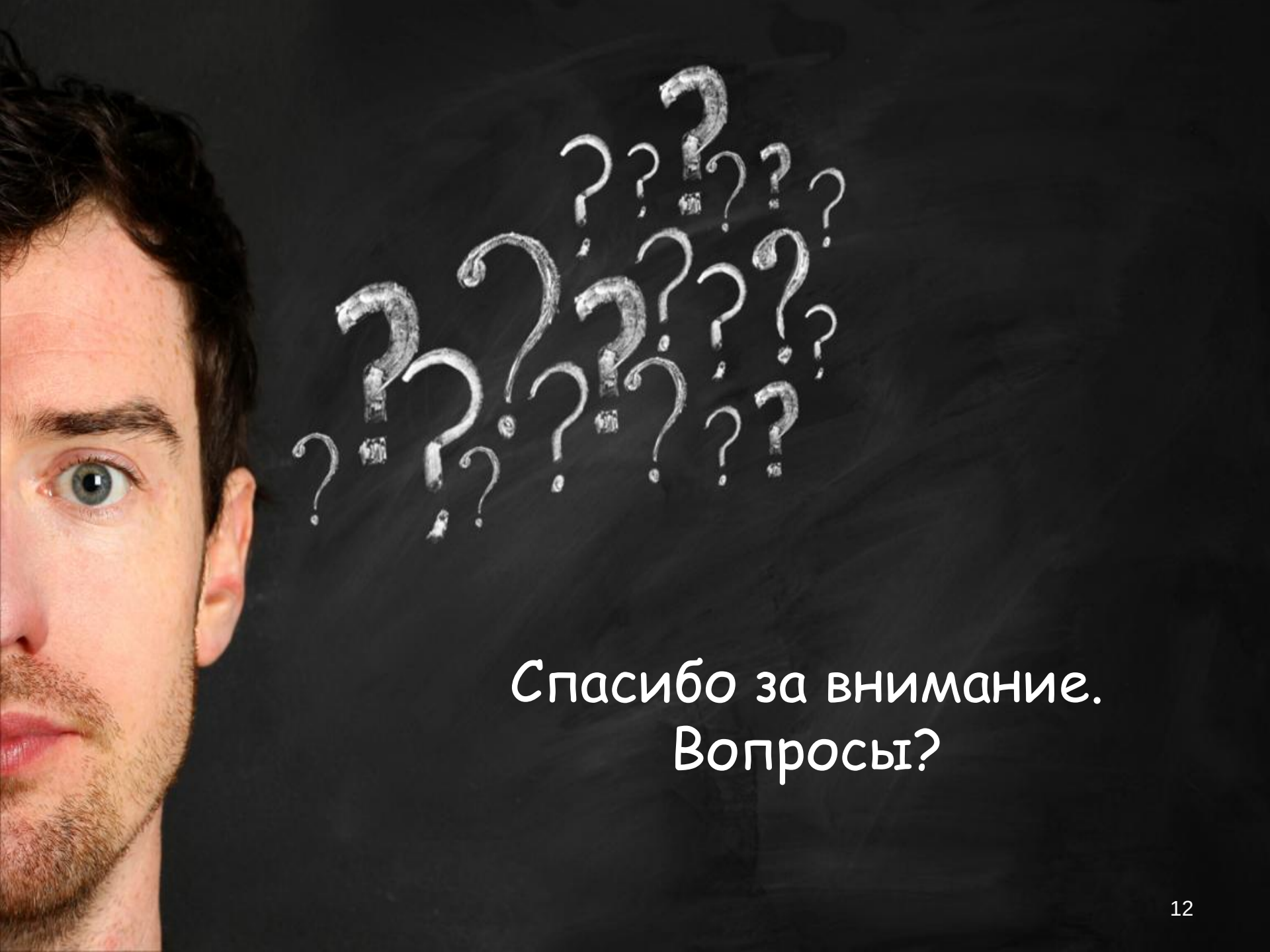
- До 70% мирового трафика передаётся в шифрованном виде (HTTPS)
- Большая часть мирового трафика на данный момент – передача потокового видео
- Для обеспечения QoS и QoE со стороны провайдера интернета и других сетевых операторов требуется оценивать некоторые характеристики:
  - Качество передаваемого видео (360, 480, 720,…)
  - Факты автоматического или ручного переключения
  - Факты пауз в просмотре из-за буферизации
  - …
- Для решения задачи часто используется машинное обучение, показавшее свою эффективность на сходных задачах классификации зашифрованного трафика по приложениям
- Также требуется анализировать новые технологии, использующиеся для адаптивной передачи потокового видео (например DASH) для понимания особенностей сетевого взаимодействия

Программный комплекс Protosphere



Основные направления работ

- Методы и алгоритмы выделения и анализа данных
  - Языки описания протоколов
  - Выявление и анализ P2P-трафика
  - Работа с прокси-серверами
  - Алгоритмы классификации
- Развитие инфраструктура универсальной системы анализа трафика
  - Универсальное API и компоненты («извлечение передаваемых файлов»)
  - Разработка системы автоматического масштабирования по входному потоку
  - Эффективная система хранения образцов трафика (параметрическая индексация)
- Разбор и анализ прикладных протоколов (HTTP/2, WebDav)
- Регулярные выражения, конечные автоматы, деревья, графы и их применение к задачам анализа
  - Анализ автоматов состояний протоколов
  - Построение и анализ графов взаимодействия («социальных графов»)
  - Выявление цепочек событий («заражение», «запуск бекдора», «утечка данных»)
- Прикладной криптоанализ (шифрованный трафик)
  - Сертификаты и ЭЦП
  - Особенности алгоритмов обмена ключами
  - Обнаружение и идентификация туннелей
  - Поведенческий анализ



Спасибо за внимание.
Вопросы?