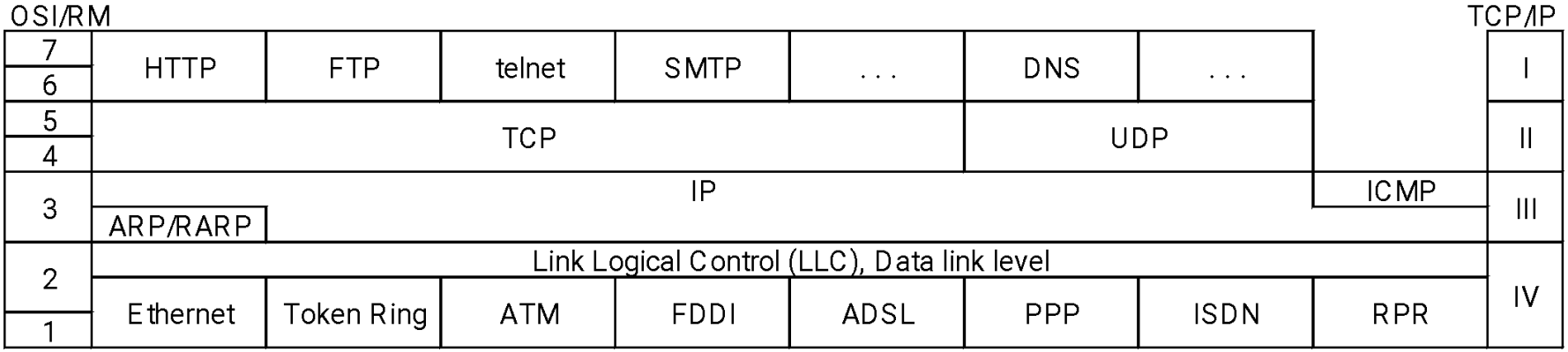
Протокол TCP/IP

OSI и TCP/IP



Стек **TCP/IP** (Transmission Control Protocol / Internet Protocol, протокол управления передачей/межсетевой протокол) в отличие от **OSI** содержит всего 4 уровня: I – прикладной, II – транспортный, III – межсетевой, IV – физический (физического интерфейса). Все они в той или степени соответствуют уровням идеальной модели, т. е. выполняют похожие функции.

Системы адресации

Адресации:

1. Прикладной уровень (служба DNS, имя компьютеров в рабочих группах Windows, др. системы символьной адресации), адресация в глобальных и локальных сетях
2. сетевой уровень (IP, IPX адреса), адресация в глобальных сетях
3. канальный уровень (MAC адрес), адресация в локальных сетях

Службы:

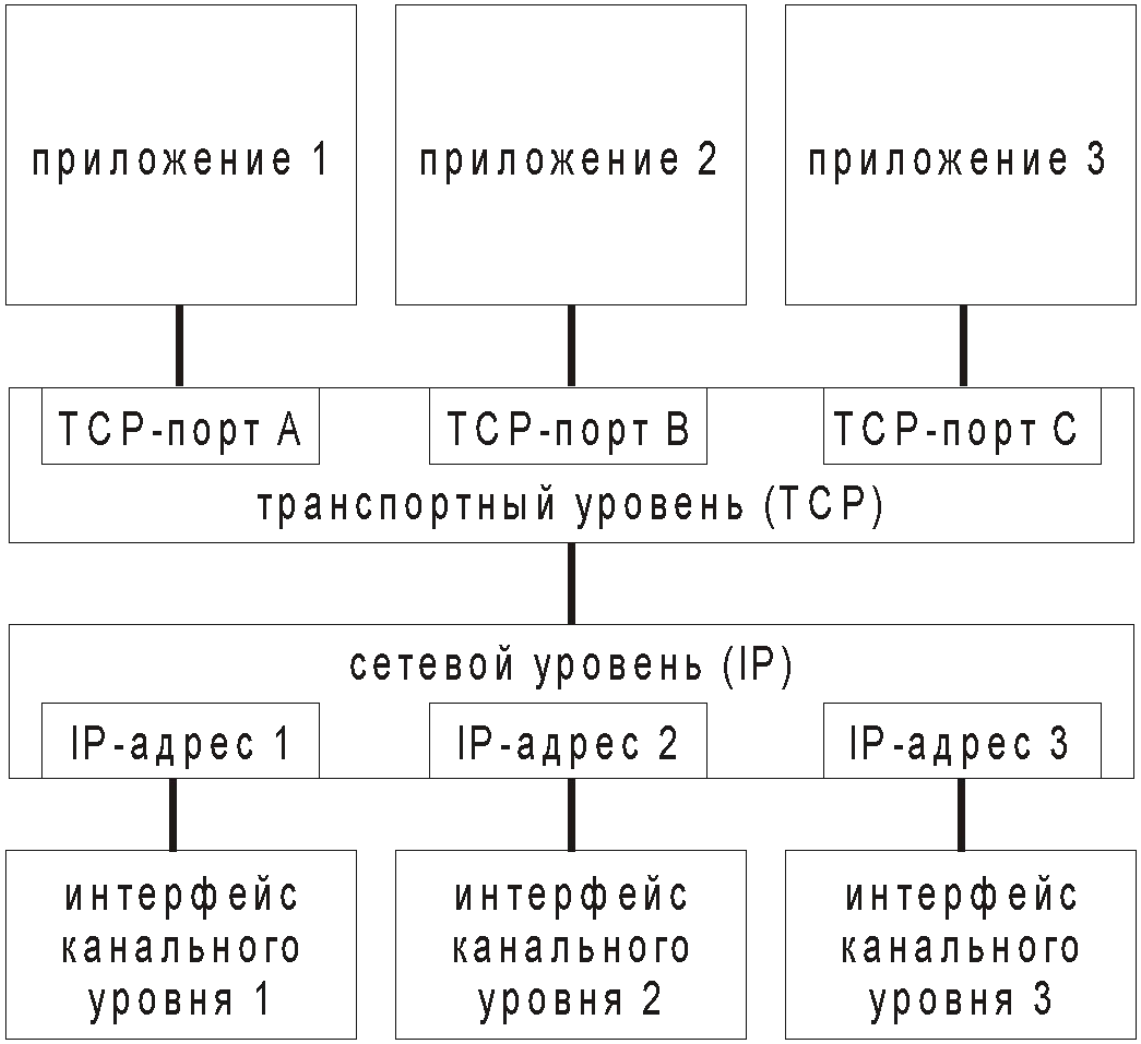
ARP/RARP (Address Resolution Protocol) – 2-3
DNS (Domain Name Service) – 1-2
Возможны службы для связи систем адресации 1-3

Типы адресов

Типы адресов

- **физический** (например, MAC-адрес 00-0b-6a-85-b6-41, ATM адрес NSAP, глобальный адрес X.25, логический адрес канального уровня)
- **сетевой** (например, IP-адрес, две части: номер сети и номер интерфейса в этой сети). Узел может иметь несколько IP адресов по количеству сетей, к которым подключен. Одному физическому интерфейсу может быть приписано несколько IP адресов, или, наоборот, одному адресу сетевого уровня соответствует несколько адресов канального уровня, но чаще всего бывает соответствие MAC адрес - IP адрес (службы ARP и RARP), например, маршрутизатор обычно имеет несколько сетевых интерфейсов с парами MAC-адрес - IP адрес.
- **логический символьный** (например, DNS-имя). Данная адресация соответствует прикладному уровню модели OSI/RM. Символьные логические адреса введены для удобства пользования глобальными адресами. Одному символьному имени может соответствовать несколько адресов сетевого уровня (например, распределенная структура серверов altavista.com) или одному IP адресу может соответствовать несколько символьных (для создания виртуальных серверов, названий веб-сайтов).

Мультиплексирование



Классы IP адресов

Класс	Структура 32-битного IP адреса		Диапазон сетей	
Класс A	0 № сети	№ хоста	1.0.0.0	126.0.0.0
Класс B	10 № сети	№ хоста	128.0.0.0	191.255.0.0
Класс C	110 № сети	№ хоста	192.0.0.0	223.255.255.0
Класс D	1110 групповой адрес		224.0.0.0	239.255.255.255
Класс E	11110 зарезервирован		240.0.0.0	247.255.255.255

IP адреса записываются в десятично-точечной нотации, каждый байт (значения в диапазоне 0-255) отделяется от соседнего точкой. Всего в Интернете возможно существование менее 2^{32} хостов (сетевых интерфейсов). Выделение хоста в сети позволяет придерживаться четкой двухуровневой иерархической структуры.

На данный момент существует бесклассовая система адресации (Classless Internet Domain Routing), характеризуемая любой длиной номера сети и хоста в пределах 32 бит.

Соглашение о спец. IP адресах

- Весь адрес состоит из 0 (0.0.0.0) - адрес данного узла, разрешается только при загрузке системы, не может быть адресом назначения.
- Поле адреса сети = 0 (например, 0.0.0.134) - узел 134 принадлежит данной сети.
- Весь адрес состоит из 1 (255.255.255.255) - ограниченное в пределах локальной сети (на канальном уровне, до первого маршрутизатора) широковещание (не может быть адресом отправителя).
- В поле хоста все 1 (например, 135.202.255.255 для сети класса B) - то широковещание в конкретной сети (не может быть адресом отправителя).
- 127.xx.xx.xx (например, 127.0.0.1) - localhost, loopback, обратная связь (никогда не передается в сеть, используется для тестирования стека TCP/IP на данном компьютере).
- Закрытые сети (синонимы: частная сеть, сеть интранет, нереальные IP адреса, серые IP адреса, нетранслируемые в Интернет IP адреса) - для соответствующих классов сетей IP адреса в следующих диапазонах:
  10.0.0.0 - 10.255.255.255
  172.16.0.0 - 172.31.255.255
  192.168.0.0 - 192.168.255.255

Маска IP адреса

Назначение маски IP адреса - отделять часть, отвечающую за номер сети от части, идентифицирующей номер хоста в данной сети.
Использование: маршрутизация и ограниченное широковещание.
Маска IP - это неразрывный последовательный бинарный ряд логических 1, оканчивающийся неразрывным рядом 0 общей длиной 32 бита.
Например,
маска IP адреса класса A: 11111111 00000000 00000000 00000000
маска IP адреса класса B: 11111111 11111111 00000000 00000000
маска IP адреса класса C: 11111111 11111111 11111111 00000000
маска длиной в 22 бита: 11111111 11111111 11111100 00000000
Иногда значение маски пишется справа после IP адреса через слеш и обозначает битовую длину части адреса, отвечающего за IP сеть, иногда в виде IP адреса, например: 134.171.0.14/25 или 255.255.255.128 – маска в 25 бит.

Выделением IP адресов в глобальном адресном пространстве ведает InterNIC (Network Information Center), в России - РосНИИРОС (Российский научно-исследовательский институт развития общественных сетей).

Использование маски

Как определить номер узла в сети, номер сети, а также адрес ограниченного в пределах данной сети широковещания?

Для этого необходимо наложить побитно маску сети на заданный IP адрес и проделать некоторые арифметические действия.

Пример 1.

Дано:

IP адрес 192.168.31.240, сеть класса C (маска в 24 бита).

Найти:

номер сети, номер узла в сети, а также адрес ограниченного в пределах данной сети широковещания.

IP адрес 11000000 10101000 00011111 11110000
маска 11111111 11111111 11111111 00000000 (255.255.255.0)

номер сети 11000000 10101000 00011111 00000000 (192.168.31.0)
номер хоста 00000000 00000000 00000000 11110000 (0.0.0.240)
адрес широковещ. 11000000 10101000 00011111 11111111 (192.168.31.255)

Использование маски

Пример 2.
IP адрес 12.200.17.242, сеть с маской в **23** бита.
IP адрес 00001100 11001000 00010001 11110010
маска 11111111 11111111 11111110 00000000 (255.255.254.0)

номер сети 00001100 11001000 00010000 00000000 (12.200.16.0)
номер хоста 00000000 00000000 00000001 11110010 (0.0.1.242)
адрес широковещ. 00001100 11001000 00010001 11111111 (12.200.17.255)

Пример 3.
IP адрес 12.200.17.242, сеть с маской в **29** бит.
IP адрес 00001100 11001000 00010001 11110010
маска 11111111 11111111 11111111 11111000 (255.255.255.248)

номер сети 00001100 11001000 00010001 11110000 (12.200.17.240)
номер хоста 00000000 00000000 00000000 00000010 (0.0.0.2)
адрес широковещ. 00001100 11001000 00010001 11110111 (12.200.17.247)

ARP, RARP

Отображение физических адресов на IP-адреса осуществляется при помощи протоколов **ARP** (Address Resolution Protocol) и **RARP** (Reversed ARP).
Сетевой IP адрес не связан с MAC адресом, как это сделано в IPX.
ARP: широковещательный запрос требуемого MAC адреса по известному IP адресу. ARP таблица (arp -a). RARP используется при старте бездисковых станций.
Формат ARP/RARP пакета (инкапсулируется в кадр канального уровня)

<table>
<tr><td colspan="2">Тип сети (1 для Eth)</td><td>Тип протокола (0080h)</td></tr>
<tr><td>Длина лок. адреса</td><td>Длина сетев. адреса</td><td>Операция (ARP=1, RARP=2)</td></tr>
<tr><td colspan="3">Локальный адрес отправителя (байты 0-3)</td></tr>
<tr><td colspan="2">Локальный адрес отправителя (4-5)</td><td>IP адрес отправителя (0-1)</td></tr>
<tr><td colspan="2">IP адрес отправителя (2-3)</td><td>Искомый локальный адрес (0-1)</td></tr>
<tr><td colspan="3">Искомый локальный адрес (байты 2-5)</td></tr>
<tr><td colspan="3">Искомый IP адрес (байты 0-3)</td></tr>
</table>

При ARP запросе поле "искомый MAC адрес" оставляют незаполненным. Значение этого поля заполняется узлом, опознавшим свой IP.

DNS

Распределенная база данных доменных имен поддерживается службой **Domain Name Service**. DNS обеспечивает иерархическую систему имен для идентификации узлов в сети Internet.

Два вида запросов в DNS сервера: прямой (по доменному имени ищется IP адрес) и обратный (доменное имя по IP адресу). Прямой поиск необходим при обычном Интернет-серфинге, когда браузер должен организовывать http сеансы связи с веб-серверами, IP адреса которых изначально неизвестны. Поиск в обратной зоне востребован некоторыми службами, например в ходе smtp связи (чаще всего с целью идентификации и примитивной защиты от взлома).

Если DNServer не знает ответа на вопрос, он пересылает запрос в домен верхнего уровня (увеличивая собственный кэш возвратившимся корректным ответом). Корень базы данных управляется центром Internet Network Information Center, в котором определены домены верхних уровней (com, gov, net, edu, mil, org, biz, info, географические домены).

Выделение доменного адреса и разделение на поддомены обеспечивается текущими владельцами доменных имен. Выделение доменного имени может быть бесплатной процедурой, если у обладателя доменного имени нет права коммерческого использования.

NSLOOKUP - программа общения с сервером DNS.

DHCP

С помощью протокола **DHCP** (Dynamic Host Configuration Protocol) автоматизирован процесс назначения IP-адресов узлам сети.
Различают статические (заранее выделенные) и динамические IP адреса. В ходе DHCP сеанса связи проходит договор не только о присвоении IP адреса данному сетевому интерфейсу, но и посылка дополнительной информации о конфигурации сети (например, адреса шлюза, маски сети, адресов прокси-серверов).
Применяется в мобильных сетях и сетях с нехваткой "реальных" (транслируемых в Интернет) адресов, при осуществлении модемного доступа к провайдеру интернет услуг.
DHCP общение происходит по архитектуре "клиент-сервер". Клиент посылает широковещательный запрос и все DHCP сервера (у каждого свой диапазон IP адресов) посылают в ответ свои конфигурационные предложения об IP адресе. После выбора хост отсылает подтверждение приема только конкретному серверу.
Существует проблема в сотрудничестве DNS и DHCP в случае динамической раздачи "реальных" IP адресов - надо постоянно обновлять DNS таблицы. Поэтому на статические сервисы, видимые из Интернет, стараются назначать "реальные" IP адреса.

Формат IP пакета

IP (Internet Protocol) пакет инкапсулируется в кадр канального уровня, чаще всего его заголовки являются вложенными в кадр IEEE 802.2.

<table>
<tr><td>версия</td><td>длина</td><td>тип сервиса</td><td colspan="2">общая длина пакета в байтах</td></tr>
<tr><td colspan="3">идентификация (для всех фрагментов одинаковое)</td><td>флаги (3 бита)</td><td>смещение фрагмента</td></tr>
<tr><td colspan="2">время жизни</td><td>протокол</td><td colspan="2">FCS заголовка</td></tr>
<tr><td colspan="5">IP адрес отправителя</td></tr>
<tr><td colspan="5">IP адрес получателя</td></tr>
<tr><td colspan="3">опции IP (если есть)</td><td colspan="2">поле заполнения до 32 бит</td></tr>
<tr><td colspan="5">данные верхних уровней</td></tr>
</table>

- Версия (IPv4)
- длина заголовка в 32 бит. словах
- тип сервиса (для интеллектуальных маршрутизаторов, PPPDTRxx, P - приоритет (для будущего), D,T,R - запрашиваются мин. задержки, макс. пропускная способность, макс. надежность)

Поля IP пакета

- Флаги Do not Fragment - DF, More Fragments - MF - еще фрагменты. (Использование - для определения MTU - Maximal Transfer Unit).
- Time to live – время жизни пакета в секундах. Это время уменьшается на количество секунд задержки на каждом маршрутизаторе или на 1 при любом переходе через маршрутизатор. Поле TTL введено для устранения бесконечного блуждания пакетов по Сети (например, в случае неправильной конфигурации маршрутизаторов и возникновения логических колец).
- Опции IP (если есть) - для тестирования или отладки сети (например, запись маршрута или обязательное прохождение по маршруту).

Минимальный размер заголовков IP уровня - 20 байт.

UDP

Заголовки и данные **UDP** (User Datagram Protocol) уровня инкапсулируются в поле данных IP уровня. UDP - протокол негарантированной доставки данных (транспортный и сеансовый уровни модели OSI/RM).

Заголовок IP (≥20 байт)	Заголовок UDP (8 байт)	Данные UDP

UDP используется для отсылки данных некритичных к потере информации приложений (DNS запросы-ответы, ICQ, TFTP, игровые сервисы типа Quake). Также UDP почти всегда используется для рассылки групповых IP датаграмм. Некоторые IP адреса класса D статически закреплены за разными сервисами, например 224.0.0.1 означает "все системы в этой подсети", а 224.0.0.2 - "все маршрутизаторы в этой подсети". Групповой адрес 224.0.1.1 предназначен для сетевого протокола времени (NTP - Network Time Protocol), а 224.0.0.9 для RIP-2. В случае групповой рассылки датаграмм с использованием адресации класса D три младшие байта IP адреса также записываются в три младшие байта адреса назначения кадра групповой рассылки канального уровня (для Ethernet).

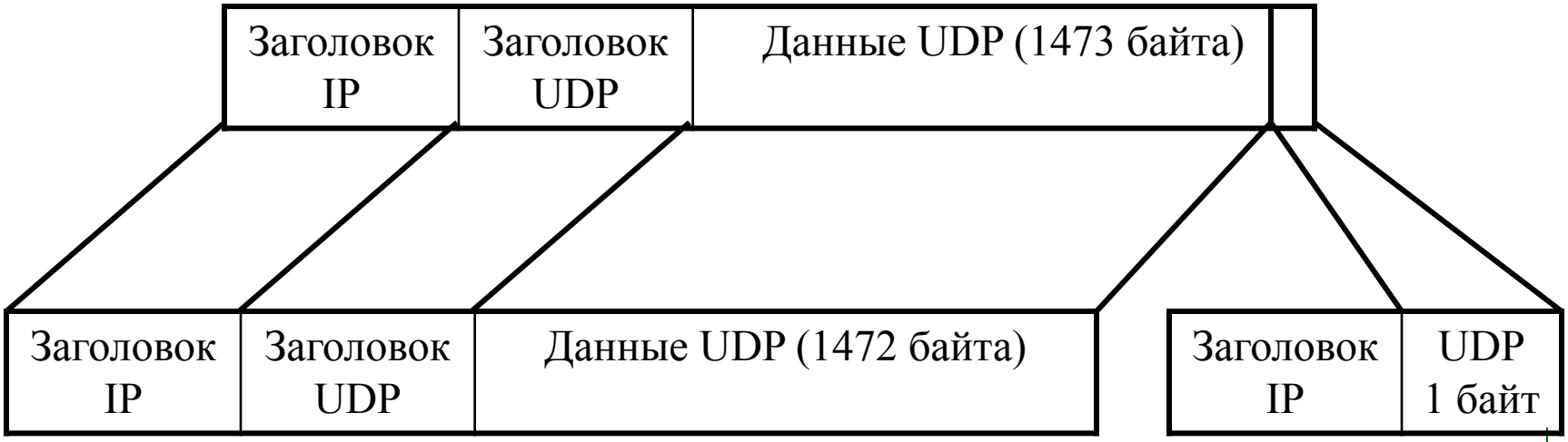
Формат UDP заголовка

Номер порта источника (16 бит)	Номер порта назначения (16 бит)
Длина UDP пакета (16 бит)	Контрольная сумма UDP (16 бит)

Количество портов источника и назначения ограничены 16-ю битами (всего 65536 портов).
Порты разделяют на именованные (закрепленные соответствующими RFC за определенными сервисами) и неименованные.
Т.к. контрольная сумма в заголовках IP уровня охватывает только заголовок, на TCP и UDP уровнях необходимо контролировать качество самих переданных данных.

IP фрагментация

Приложениям, которые пользуются UDP для отправки данных, нет необходимости заботиться о размере получившейся в результате IP датаграммы (лишь бы она не выходила за пределы 64кб, максимального размера). Если она по размеру больше, чем MTU для данной сети, IP датаграмма будет фрагментирована. На рисунке приведен пример фрагментации поверх Ethernet.



TCP

Заголовки и данные **TCP** (Transmission Control Protocol) уровня инкапсулируются в поле данных IP уровня, т.е. в IP датаграмму. TCP - протокол гарантированной доставки данных по предустановленному виртуальному соединению (транспортный и сеансовый уровни модели OSI/RM).

Заголовок IP (≥20 байт)	Заголовок TCP (≥20 байт)	Данные TCP

Единицей данных протокола TCP является сегмент. Оба участника соединения должны договориться о максимальном размере сегмента, который они будут использовать. Этот размер выбирается таким образом, чтобы при упаковке сегмента в IP-пакет он помещался туда целиком, то есть максимальный размер сегмента не должен превосходить максимального размера поля данных IP-пакета. Максимальный размер сегмента не должен превышать минимальное значение на множестве всех MTU промежуточных IP сетей.
TCP строит пакеты, упаковывая их в сегменты, устанавливает тайм-ауты в момент отправки, подтверждает принятые данные, меняет их порядок в случае хаотического прибытия (вследствие различных путей датаграмм), отбрасывает дублированные данные, осуществляет контроль потока данных, рассчитывает и проверяет контрольную сумму.

Формат TCP заголовков

<table>
<tr><td colspan="5">номер порта источника (16 бит)</td><td colspan="4">номер порта назначения (16 бит)</td></tr>
<tr><td colspan="9">номер последовательности (32 бита)</td></tr>
<tr><td colspan="9">номер подтверждения (32 бита)</td></tr>
<tr><td>4 бита длина заголовка</td><td>резерв 6 бит</td><td>URG</td><td>ACK</td><td>PSH</td><td>RST</td><td>SYN</td><td>FIN</td><td>размер окна (16 бит)</td></tr>
<tr><td colspan="5">контрольная сумма (16 бит)</td><td colspan="4">указатель срочности (16 бит)</td></tr>
<tr><td colspan="9">опции (если есть)</td></tr>
<tr><td colspan="9">данные (если есть)</td></tr>
</table>

- Номер последовательности (sequence number) идентифицирует количество байт в переданном потоке в одном направлении. При установлении нового соединения значения этого поля содержит исходный номер последовательности (выбирается псевдослучайным образом).
- Поле номер подтверждения содержит номер последнего успешно принятого байта +1.

Формат TCP заголовков

- Длина заголовка выражается в 4х байтовых словах (максимальная длина TCP заголовка - 60 байт).
- Битовые флаги:
  - URG - флаг срочности (запрет ожидания заполнения исходящего буфера при передаче), используется совместно с указателем срочности (смещением, складываемым с номером последовательности). Флаг используется, например, при нажатии CTRL+C в режиме telnet.
  - ACK - указатель подтверждения приема
  - PSH - получатель должен передать данные приложению как можно быстрее (используется очень часто для уменьшения времени передачи информации).
  - RST - сброс соединения.
  - SYN - сигнал установления соединения.
  - FIN - отправитель заканчивает отсылку данных.
- Контроль потока данных на каждой стороне TCP соединения производится с использованием окна (0-65535 байт). Это количество байт, начинающееся с указанного в поле номера подтверждения, которое приложение собирается принять.

Режим promiscuous

В обычном режиме функционирования сетевого интерфейса при получении кадра данные (46–1500 байтов) будут переданы обработчику верхнего уровня только в случаях, если адрес назначения, установленный в поле DA, широковещательный, либо он совпадет с уникальным **MAC**-адресом принимающей станции.

Однако каждый адаптер Ethernet может быть переведен в режим, в котором будут обрабатываться все кадры, поступающие из среды передачи. На английском языке такой режим носит название "promiscuous", что переводится как "безразличный" или "неразборчивый". Этим свойством сетевых адаптеров можно пользоваться, например, для создания программных анализаторов сетевого трафика (**tcpdump**).