UP Down Next INDEX Поиск

UP: 4.4 Интернет
Down: 4.4.1.1 Адресация IPv6
Next: 4.4.2 Протокол UDP

4.4.1 IP-протокол

Семенов Ю.А. (ИТЭФ-МФТИ)
Semenov Yu (ITEP-MIPT)

Номер раздела	Название раздела	Объем в страницах	Объем в кбайт
4.4.1.1	Адресация IPv6	57	485
4.4.1.2	IP-туннели	3	87
4.4.1.3	Протокол туннелей на сетевом уровне L2 (L2TP)	54	48
Итого			

ToS в IP-протоколе
Замещение ToS на DSCP
Коды протоколов Интернет
Опции IP-протокола
Распределение протоколов по уровням

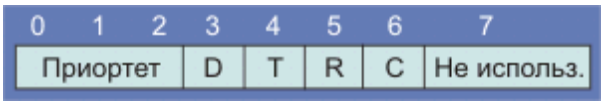
В Интернет используется много различных типов пакетов, но один из основных - IP-пакет (RFC-791), именно он вкладывается в кадр Ethernet и именно в него вкладываются пакеты UDP, TCP. IP-протокол предлагает ненадежную транспортную среду. Ненадежную в том смысле, что не существует гарантии благополучной доставки IP-дейтограммы. Алгоритм доставки в рамках данного протокола предельно прост: при ошибке дейтограмма выбрасывается, а отправителю посылается соответствующее ICMP-сообщение (или не посылается ничего). Обеспечение же надежности возлагается на более высокий уровень (UDP или TCP). Формат IP-пакетов показан на рисунке 4.4.1.1.

Рис. 4.4.1.1. Формат дейтограммы Интернет

Поле *версия* характеризует версию IP-протокола (например, 4 или 6). Формат пакета определяется программой и, вообще говоря, может быть разным для разных значений поля версия. Только размер и положение этого поля незыблемы. Поэтому в случае изменений длины IP-адреса слишком тяжелых последствий это не вызовет. Понятно также, что значение поля версия во избежании непредсказуемых последствий должно контролироваться программой. *HLEN* - длина заголовка, измеряемая в 32-разрядных словах, обычно заголовок содержит 20 октетов (HLEN=5, без опций и заполнителя). **Заголовок для IPv6 имеет размер в два раза больше, чем для IPv4.** Поле *полная длина* определяет полную длину IP-дейтограммы (до 65535 октетов), включая заголовок и данные.

Заголовок IPv4 может расширяться от 20 до максимум 60 байт, но эта опция редко используется, так как влечет ухудшение рабочих характеристик и часто административно запрещена из соображений безопасности.

Одно-октетное поле *тип сервиса* (TOS - type of service) характеризует то, как должна обрабатываться дейтограмма. Это поле делится на 6 субполей:



Субполе *Приоритет* предоставляет возможность присвоить код приоритета каждой дейтограмме. Значения приоритетов приведены в таблице (в настоящее время это поле не используется).

0 Обычный уровень
1 Приоритетный
2 Немедленный
3 Срочный

4 Экстренный
5 ceitic/ecp
6 Межсетевое управление
7 Сетевое управление

Формат поля TOS определен в документе RFC-1349. Биты C, D, T и R характеризуют пожелание относительно способа доставки дейтограммы. Так D=1 требует минимальной задержки, T=1 - высокую пропускную способность, R=1 - высокую надежность, а C=1 - низкую стоимость. TOS играет важную роль в маршрутизации пакетов. Интернет не гарантирует запрашиваемый TOS, но многие маршрутизаторы учитывают эти запросы при выборе маршрута (протоколы OSPF и IGRP). В таблице 4.4.1.1 приведены рекомендуемые значения TOS.

ToS в IP-протоколе

Таблица 4.4.1.1. Значения TOS для разных протоколов

Процедура	Минимал. задержка	Максим. пропускная способность	Максим. надежность	Минимал. стоимость	Код TOS
FTP управление, FTP данные	1	0	0	0	0x10
	0	1	0	0	0x08
TFTP	1	0	0	0	0x10
DNS, UDP TCP	1	0	0	0	0x00
	0	0	0	0	0x10
	0	0	0	0	0x00
telnet	1	0	0	0	0x10
ICMP	0	0	0	0	0x00
IGP	0	0	1	0	0x04
SMTP управление SMTP данные	1	0	0	0	0x10
	0	1	0	0	0x08
SNMP	0	0	1	0	0x04

NNTP	0	0	0	1	0x02

Только один бит из четырех в TOS может принимать значение 1. Значения по умолчанию равны нулю. Большинство из рекомендаций самоочевидны. Так при telnet наибольшую важность имеет время отклика, а для SNMP (управление сетью) - надежность.

Замещение ToS на DSCP

До середины 90-х годов поле TOS в большинстве реализаций игнорировалось. Но после начала разработок средств обеспечения качества обслуживания (QoS) внимание к этому возросло. Появилось предложение замены поля TOS на поле DSCP (Differenciated Services Code Point), которое также имеет 8 бит (см. RFC-2474). Смотри рис. 4.4.1.1а. Биты CU пока не определены. Иногда это поле называется байтом DS (Differentiated Services).

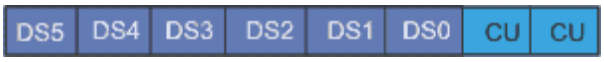


Рис. 4.4.1.1а. Формат поля DSCP.

Биты DS0-DS5 определяют селектор класса. Значения этого кода представлены в таблице ниже. Стандартным значением DSCP по умолчанию является 000000.

Селектор класса	**DSCP**
Приоритет 1	001000
Приоритет 2	010000
Приоритет 3	011000
Приоритет 4	100000
Приоритет 5	101000
Приоритет 6	110000
Приоритет 7	111000

На базе DSCP разработана технология "пошагового поведения" PHB (per

Hop Behavior). В рамках этой политики определяются коды DSCP внутри классов. Например, для политики немедленной переадресации EF рекомендуемое значение DSCP=101110. Эта политика соответствует наиболее высокому уровню обслуживания.

Маршрут транспортировки IP-дейтограммы нельзя знать заранее, это связано с поэтапным (по-шаговом) принятием решения о пути каждого пакета. Это свойство маршрутизации обусловлено тем, что IP является протоколом передачи данных без установления соединения.

Поля *идентификатор*, *флаги* (3 бита) и *указатель фрагмента* (fragment offset) управляют процессом фрагментации и последующей "сборки" дейтограммы. *Идентификатор* представляет собой уникальный код дейтограммы, позволяющий идентифицировать принадлежность фрагментов и исключить ошибки при "сборке" дейтограмм. Значение идентификатора определяется верхним протокольным уровнем. Поле *указатель фрагмента* указывает место, соответствующее этому фрагменту в дейтограмме. Это положение измеряется в единицах, равных 8 октетам (64 бита). Первому фрагменту соответствует указатель равный нулю.

Бит 0 поля *флаги* является резервным, бит 1 служит для управления фрагментацией пакетов (0 - фрагментация разрешена; 1 - запрещена), бит 2 определяет, является ли данный фрагмент последним (0 - последний фрагмент; 1 - следует ожидать продолжения). Поле *время жизни* (**TTL** - time to live) задает время жизни дейтограммы в секундах, т.е. предельно допустимое время пребывания дейтограммы в системе. При каждой обработке дейтограммы, например в маршрутизаторе, это время уменьшается в соответствии со временем пребывания в данном устройстве или согласно протоколу обработки. Если TTL=0, дейтограмма из системы удаляется. Во многих реализациях TTL измеряется в числе шагов, в этом случае каждый маршрутизатор выполняет операцию TTL=TTL-1. TTL помогает предотвратить зацикливание пакетов. Поле *протокол* аналогично полю *тип* в Ethernet-кадре и определяет структуру поля *данные* (см. табл. 4.4.1.2).

Поле TTL относится к числу переменных полей заголовка. При прохождении через маршрутизатор над содержимым этого поля производится операция **TTL=TTL-1**, при этом должна быть пересчитана контрольная сумма. И, если TTL=0, дейтограммы отбрасывается.

Поле *контрольная сумма заголовка* вычисляется с использованием операций сложения 16-разрядных слов заголовка по модулю 1. Сама контрольная сумма является дополнением по модулю один полученного

результата сложения. Обратите внимание, здесь осуществляется контрольное суммирование заголовка, а не всей дейтограммы. Поле *опции* не обязательно присутствует в каждой дейтограмме. Размер поля *опции* зависит от того, какие опции применены. Если используется несколько опций, они записываются подряд без каких-либо разделителей. Каждая опция содержит один октет кода опции, за которым может следовать октет длины и серия октетов данных. Если место, занятое опциями, не кратно 4 октетам, используется заполнитель. Структура октета кода опции отражена на рис. 4.4.1.2.

Коды протоколов Интернет

Таблица 4.4.1.2. Коды протоколов Интернет

Код протокола Интернет	Сокращенное название протокола	Описание
0	-	Зарезервировано
1	ICMP	Протокол контрольных сообщений [rfc792]
2	IGMP	Групповой протокол управления [rfc1112]
3	GGP	Протокол маршрутизатор-маршрутизатор [RFC-823]
4	IP	IP поверх IP (инкапсуляция/туннели)
5	ST	Поток [rfc1190]
6	TCP	Протокол управления передачей [RFC-793]
7	UCL	UCL
8	EGP	Протокол внешней маршрутизации [RFC-888]
9	IGP	Протокол внутренней маршрутизации
10	BBN-MON	BBN-RCC мониторирование
11	NVP-II	Сетевой протокол для голосовой связи [RFC-741]
12	PUP	PUP
13	ARGUS	argus

14	Emcon	emcon
15	Xnet	Перекрестный сетевой отладчик [IEN158]
16	Chaos	Chaos
17	UDP	Протокол дейтограмм пользователя [RFC-768]
18	MUX	Мультиплексирование [IEN90]
19	DCN-MEAS	DCN измерительные субсистемы
20	HMP	Протокол мониторирования ЭВМ (host [RFC-869])
21	PRM	Мониторирование при передаче пакетов по радио
22	XNS-IDP	Xerox NS IDP
23	Trunk-1	Trunk-1
24	Trank-2	Trunk-2
25	Leaf-1	Leaf-1
26	Leaf-2	Leaf-2
27	RDP	Протокол для надежной передачи данных [RFC-908]
28	IRTP	Надежный TP для Интернет [RFC-938]
29	ISO-TP4	ISO транспортный класс 4 [RFC-905]
30	Netblt	Массовая передача данных [RFC-969]
31	MFE-NSP	Сетевая служба MFE
32	Merit-INP	Межузловой протокол Merit
33	SEP	Последовательный обмен
34		не определен
35	IDRP	Междоменный протокол маршрутизации
36	XTP	Xpress транспортный протокол
37	DDP	Протокол доставки дейтограмм

38	IDPR-CMTP	IDPR передача управляющих сообщений
39	TP++	TP++ транспортный протокол
40	IL	IL-транспортный протокол
41	SIP	Простой Интернет-протокол
42	SDRP	Протокол маршрутных запросов для отправителя
43	SIP-SR	SIP исходный маршрут
44	SIP-Frag	SIP-фрагмент
45	IDRP	Интер-доменный маршрутный протокол
46	RSVP	Протокол резервирования ресурсов канала
47	GRE	Общая инкапсуляция маршрутов
49	BNA	BNA
50	SIPP-ESP	SIPP ES3
52	I-NLSP	Интегрированная система безопасности сетевого уровня
53	Swipe	IP с кодированием
54	NHRP	nbma протокол определения следующего шага
55-60		не определены
61		Любой внутренний протокол ЭВМ
62	CFTP	CFTP
63		Любая локальная сеть
64	Sat-Expak	Satnet и Expak
65	MIT-Subn	Поддержка субсетей MIT
66	RVD	Удаленный виртуальный диск MIT
67	IPPC	IPPC
68		Любая распределенная файловая система

69	Sat-Mon	Мониторирование Satnet
70		не определен
71	IPCV	Базовая пакетная утилита
75	PVP	Пакетный видео-протокол
76	BRsat-Mon	Резервное мониторирование Satnet
78	Wb-mon	Мониторирование Expak
79	Wb-expak	Широкополосная версия Expak
80	ISO-IP	ISO Интернет протокол
88	IGRP	IGRP (Cisco) - внутренний протокол маршрутизации
89	OSPFIGP	OSPFIGP - внутренний протокол маршрутизации
92	MTP	Транспортный протокол мультикастинга
101-254		не определены
255		зарезервировано

Опции IP-протокола

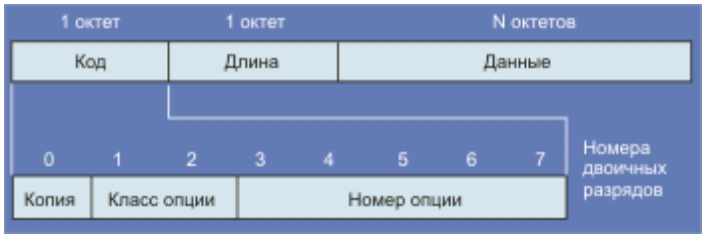


Рис. 4.4.1.2. Формат описания опций

Флаг *копия* равный 1 говорит о том, что опция должна быть скопирована во все фрагменты дейтограммы. При равенстве этого флага 0 опция копируется только в первый фрагмент. Ниже приведены значения разрядов 2-битового поля *класс опции*:

Значение поля *класс опции*	Описание
0	Дейтограмма пользователя или сетевое управление
1	Зарезервировано для будущего использования
2	Отладка и измерения (диагностика)
3	Зарезервировано для будущего использования

В таблице, которую вы найдете ниже, приведены значения классов и номеров опций.

Класс опции	Номер опции	Длина описания	Назначение
0	0	-	Конец списка опций. Используется, если опции не укладываются в поле заголовка (смотри также поле "заполнитель")
0	1	-	Никаких операций (используется для выравнивания октетов в списке опций)
0	2	11	Ограничения, связанные с секретностью (для военных приложений)
0	3	*	Свободная маршрутизация. Используется для того, чтобы направить дейтограмму по заданному маршруту
0	7	*	Запись маршрута. Используется для трассировки
0	8	4	Идентификатор потока. Устарело.
0	9	*	Жесткая маршрутизация. Используется, чтобы направить дейтограмму по заданному маршруту
2	4	*	Временная метка Интернет

* в колонке "длина" - означает - переменная.

Наибольший интерес представляют собой опции *временные метки* и *маршрутизация*. Опция *записать маршрут* (RR) создает дейтограмму, где зарезервировано место, куда каждый маршрутизатор по дороге должен записать свой IP-адрес (например, утилита traceroute). Формат опции *записать маршрут* в дейтограмме представлен ниже на рис. 4.4.1.3 (предусмотрено место для записи 9 IP-адресов, к сожаления, реализация RR не является обязательной):

Рис. 4.4.1.3 Формат опций *записать маршрут*

Поле *код* содержит номер опции (7 в данном случае). Поле *длина* определяет размер записи для опций, включая первые 3 октета. *Указатель* отмечает первую свободную позицию в списке IP-адресов (куда можно произвести запись очередного адреса). Интересную возможность представляет опция *маршрут отправителя*, которая открывает возможность посылать дейтограммы по заданному отправителем маршруту. Это позволяет исследовать различные маршруты, в том числе те, которые недоступны через узловые маршрутизаторы. Существует две формы такой маршрутизации: *Свободная маршрутизация* и *Жесткая маршрутизация* (маршрутизация отправителя). Форматы для этих опций показаны ниже:

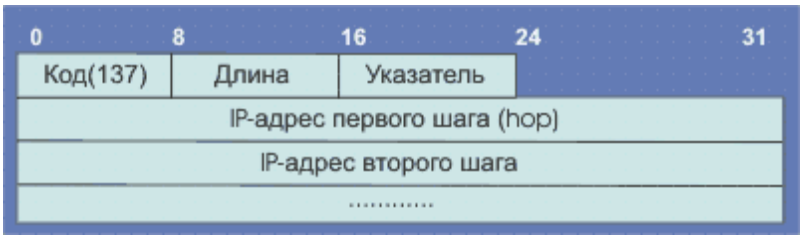


Рис. 4.4.1.3а. Формат опций маршрутизации

Жесткая маршрутизация означает, что адреса определяют точный маршрут дейтограммы. Проход от одного адреса к другому может включать только одну сеть. *Свободная маршрутизация* отличается от предшествующей возможностью прохода между двумя адресами списка более чем через одну сеть. Поле *длина* задает размер списка адресов, а *указатель* отмечает адрес очередного маршрутизатора на пути дейтограммы.

IP-слой имеет маршрутные таблицы, которые просматриваются каждый раз,

когда IP получает дейтограмму для отправки. Когда дейтограмма получается от сетевого интерфейса, IP первым делом проверяет, принадлежит ли IP-адрес места назначения к списку локальных адресов, или является широковещательным адресом. Если имеет место один из этих вариантов, дейтограмма передается программному модулю в соответствии с кодом в поле протокола. IP-процессор может быть сконфигурирован как маршрутизатор, в этом случае дейтограмма может быть переадресована в другой узел сети. Маршрутизация на IP-уровне носит пошаговый характер. IP не знает всего пути, он владеет лишь информацией - какому маршрутизатору послать дейтограмму с конкретным адресом места назначения.

Просмотр маршрутной таблицы происходит в три этапа:

1. Ищется полное соответствие адресу места назначения. В случае успеха, пакет посылается соответствующему маршрутизатору или непосредственно интерфейсу адресата. Связи точка-точка выявляются именно на этом этапе.
2. Ищется соответствие адресу сети места назначения. В случае успеха система действует также как и в предшествующем пункте. Одна запись в таблице маршрутизации соответствует всем ЭВМ, входящим в данную сеть.
3. Осуществляется поиск маршрута по умолчанию и, если он найден, дейтограмма посылается в соответствующий маршрутизатор.

Для того чтобы посмотреть, как выглядит простая маршрутная таблица, воспользуемся командой netstat -rn (ЭВМ Sun. Флаг -r выводит на экран маршрутную таблицу, а -n отображает IP-адреса в цифровой форме. С целью экономии места таблица в несколько раз сокращена).

routing tables destination	gateway	flags	refcnt	use	interface
193.124.225.72	193.124.224.60	ughd	0	61	le0
192.148.166.1	193.124.224.60	ughd	0	409	le0
193.124.226.81	193.124.224.37	ughd	0	464	le0
192.160.233.201	193.124.224.33	ughd	0	222	le0
192.148.166.234	193.124.224.60	ughd	1	3248	le0
193.124.225.66	193.124.224.60	ughd	0	774	le0
192.148.166.10	193.124.224.60	ughd	0	621	le0
192.148.166.250	193.124.224.60	ughd	0	371	le0

192.148.166.4	193.124.224.60	ughd	0	119	le0
145.249.16.20	193.124.224.60	ughd	0	130478	le0
192.102.229.14	193.124.224.33	ughd	0	13206	le0
default	193.124.224.33	ug	9	5802624	le0
193.124.224.32	193.124.224.35	u	6	1920046	le0
193.124.134.0	193.124.224.50	ugd	1	291672	le0

Колонка destination - место назначение, Default - отмечает маршрут по умолчанию; Gateway - IP-адреса портов подключения (маршрутизаторов); REFCNT (reference count) - число активных пользователей маршрута; USE - число пакетов, посланных по этому маршруту; interface - условные имена сетевых интерфейсов. Расшифровка поля FLAGS приведено ниже:

u	Маршрут работает (up).
g	Путь к маршрутизатору (gateway), если этот флаг отсутствует, адресат доступен непосредственно.
h	Маршрут к ЭВМ (host), адрес места назначения является полным адресом этой ЭВМ (адрес сети + адрес ЭВМ). Если флаг отсутствует, маршрут ведет к сети, а адрес места назначения является адресом сети.
d	Маршрут возник в результате переадресации.
m	Маршрут был модифицирован с помощью переадресации.

Опция *временные метки* работает также как и опция *запись маршрута*. Каждый маршрутизатор на пути дейтограммы делает запись в одном из полей дейтограммы (два слова по 32 разряда; смотри раздел 4.4.15). Формат этой опции отображен на рисунке 4.4.1.4.

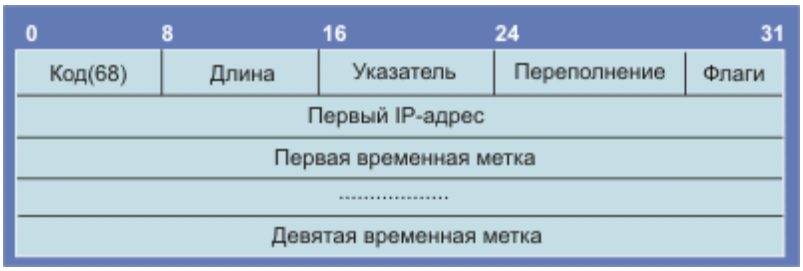


Рис. 4.4.1.4 Формат опции "временные метки"

Смысл полей ***длина*** и ***указатель*** идентичен тому, что сказано о предыдущих опциях. 4-битовое поле ***переполнение*** содержит число маршрутизаторов, которые не смогли записать временные метки из-за ограничений выделенного места в дейтограмме. Значения поля ***флаги*** задают порядок записи временных меток маршрутизаторами:

Таблица 4.4.1.3.

Значение флага	**Назначение**
0	Записать только временные метки; опустить IP-адреса.
1	Записать перед каждой временной меткой IP-адрес (как в формате на предыдущем рисунке).
3	IP-адреса задаются отправителем; маршрутизатор записывает только временные метки, если очередной IP-адрес совпадает с адресом маршрутизатора

Временные метки должны содержать время в миллисекундах, отсчитанное от начала суток. Если маршрутизатору некуда положить свою временную метку (число меток превысило 9), он инкрементирует счетчик ***переполнение***.

IPv4 имеет ограниченное место для опций (только 40 байт), и, следовательно, структура пакетов имеет ограничения для дальнейших функциональных расширений. Кроме того, маршрутизаторы, переадресуя пакеты, должны обрабатывать все имеющиеся опции.

Взаимодействие других протоколов с IP можно представить из схемы на рис. 4.4.1.5. В основании лежат протоколы, обеспечивающие обмен информацией на физическом уровне, далее следуют протоколы IP, ICMP, ARP, RARP, IGMP и протоколы маршрутизаторов. Чем выше расположен протокол, тем более высокому уровню он соответствует. Протоколы, имена которых записаны в одной и той же строке, соответствуют одному и тому же уровню. Но все разложить аккуратно по слоям невозможно - некоторые протоколы занимают промежуточное положение, что и отражено на схеме, (области таких протоколов захватывают два уровня. Здесь протоколы IP, ICMP и IGMP помещены на один уровень, для чего имеется не мало причин. Но иногда последние два протокола помещают над IP, так как их пакеты вкладываются в IP-дейтограммы. Так что деление протоколов по уровням довольно условно. На самом верху пирамиды находятся прикладные программы, хотя пользователю доступны и более низкие уровни (например,

ICMP), что также отражено на приведенном рисунке (4.4.1.5).

Распределение протоколов по уровням

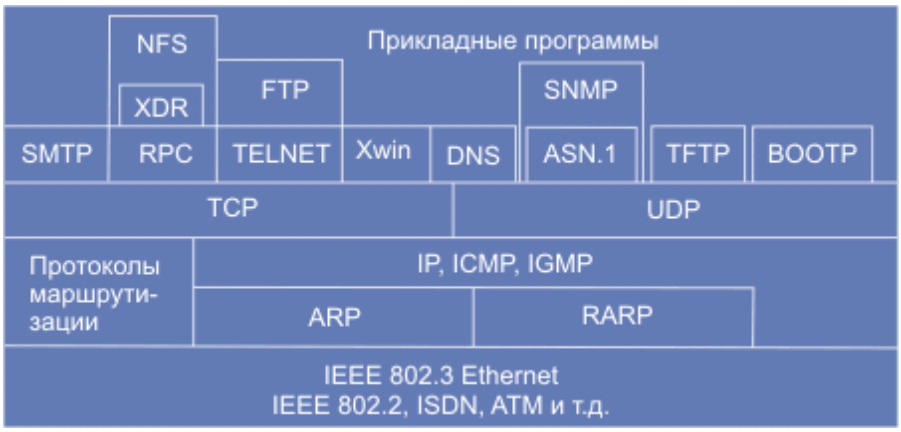


Рис. 4.4.1.5. Распределение протоколов Интернет по уровням

Интернет - это инструмент общения, средство доступа к информации и как всякий инструмент требует практики. Из вашего собственного опыта вы знаете, что можно прочесть ворох инструкций о том, как забивать гвозди, но научиться этому можно лишь на практике. Поэтому рекомендую с самого начала, читая данные тексты, чаще садитесь за терминал.

UP: 4.4 Интернет
Down: 4.4.1.1 Адресация IPv6
Next: 4.4.2 Протокол UDP