10. Специални IP адреси. NAT.

Преобразуване на IP адреси във физически. ARP vs. RARP. DHCP. ICMP.

Специални IP адреси

В рамките на IPv4 адресното пространство има адресни сегменти, които са отделени за частно (локално) използване.

RFC3330 прави "карта" на адресните сегменти за специално използване.

Специални IP адреси

127.0.0.0/8 - Internet host loopback address. Пакет се зацикля вътре в хоста. И не се появява никъде в мрежата.

169.254.0.0/16 – това е "link local" блок. Хостовете получават такива адреси по “auto-configuration”, например не може да се намери DHCP сървър.

Частни IP адреси

Те не се маршрутизират глобално, а само локално, за локални (частни) цели.

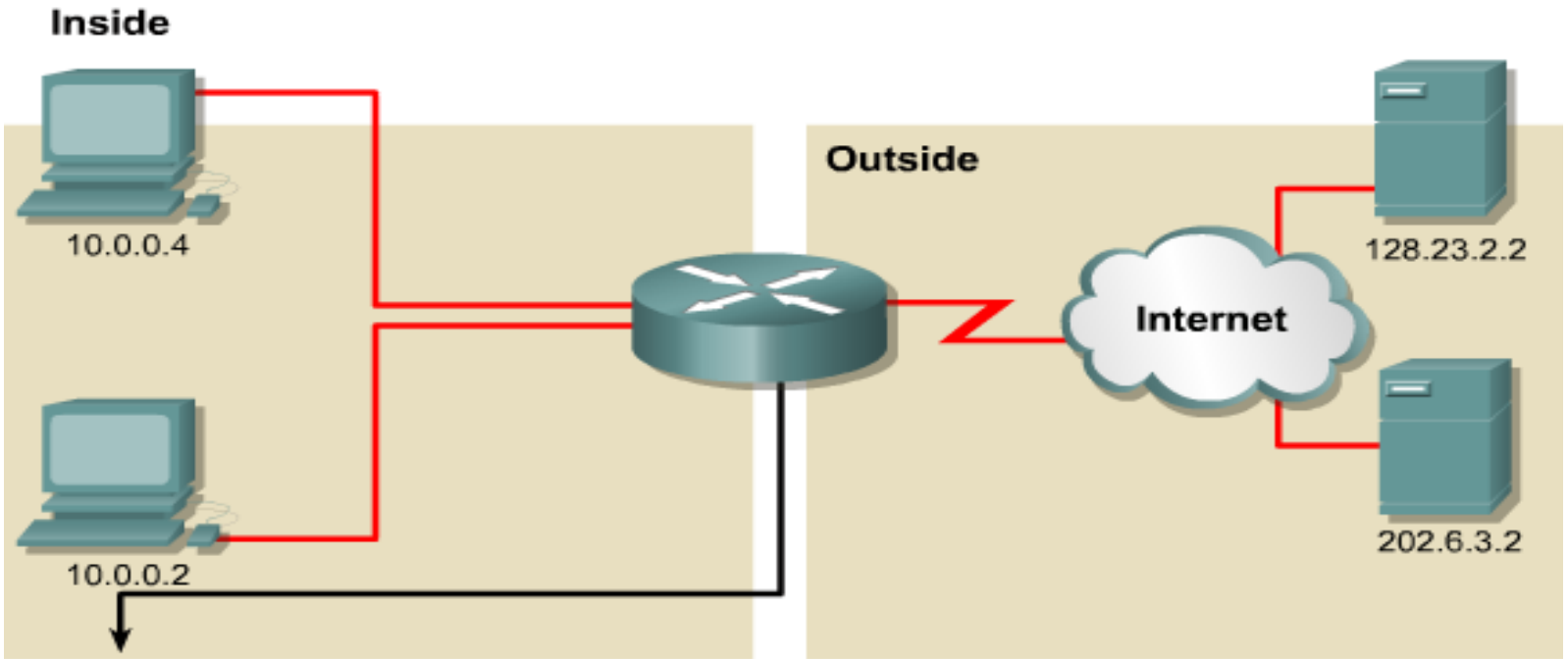
RFC3330 (първото указване е в RFC1918) указва кои от адресните пространства се използват за частни цели:

10.0.0.0/8 т.е 10.0.0.0 – 10.255.255.255

172.16.0.0/12 т.е 172.16.0.0 - 172.31.255.255

192.168.0.0/16 т.е 192.168.0.0 - 192.168. 255.255

NAT (Network Address Translation)



NAT Table with Overload

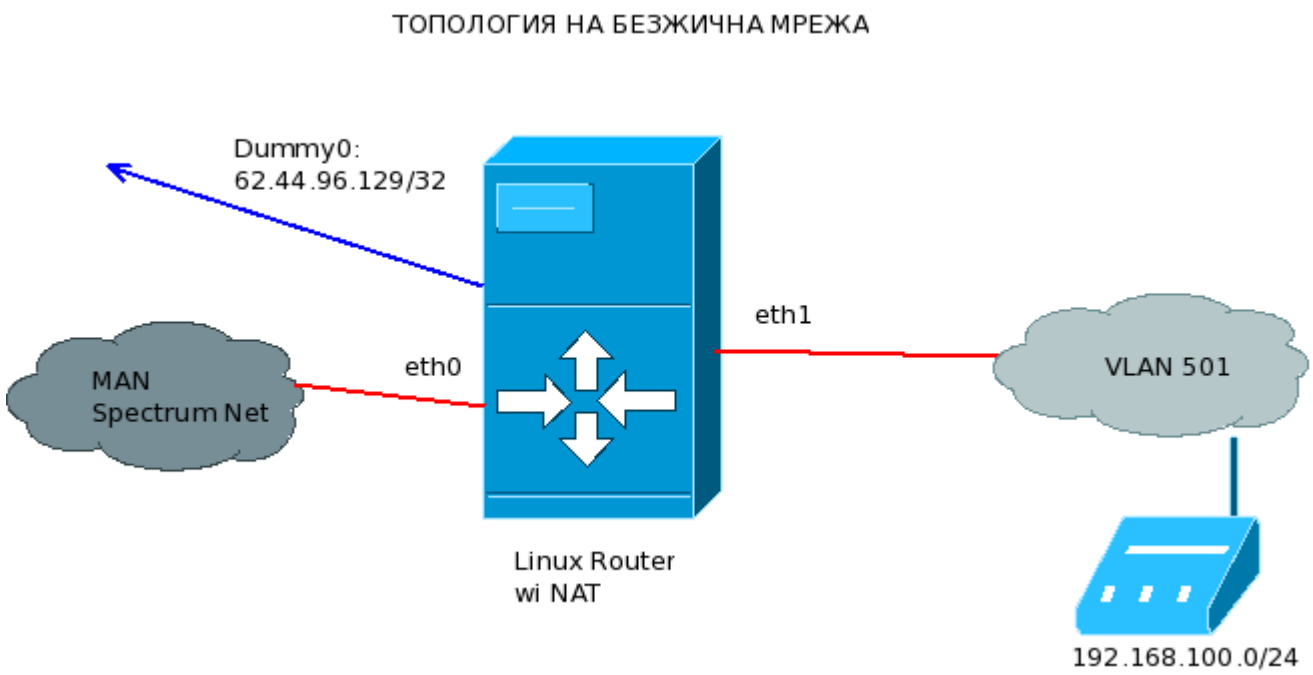
Inside Local IP Address	Inside Global IP Address	Outside Local IP Address	Outside Global Address
10.0.0.2:1331	179.9.8.80:1331	202.6.3.2:80	202.6.3.2:80
10.0.0.4:1555	179.9.8.80:1555	128.23.2.2:80	128.23.2.2:80

Statefull NAT

NAT в Linux ядрата 2.6 се описва в специални таблици в паметта - connection tracking tables.

Такъв тип NAT е statefull - специално алокирани за целта страници в паметта.

Безжична мрежа зад NAT



Безжична мрежа зад NAT (dummy0)

Публичният IP адрес за NAT да не бъде директно достъпен в никой физически сегмент. Може да бъде променен.

Създава се **dummy** интерфейс.

/etc/sysconfig/network-scripts/ifcfg-dummy0:

```
DEVICE=dummy0
IPADDR=62.44.96.129
NETMASK=255.255.255.255
ONBOOT=yes
```

ifup dummy0

Безжична мрежа зад NAT (Конфигуриране)

Чрез инструмента iptables и едноименната услуга.

Указва се правилото за NAT (като root):

```
# iptables -t nat -A POSTROUTING
-s 192.168.100.0/24 -d 0.0.0.0/0
-o eth0 -j SNAT --to 62.44.96.129
```

Всички изходящи от 192.168.100.0/24 (WiFi потребители) пакети през интерфейса eth0 се маскират към публичния адрес 62.44.96.129 (dummy0).

Безжична мрежа зад NAT (connection tracking tables)

```
# service iptables save
# chkconfig iptables on
```

```
ip_conntrack: table full, dropping packet
```

Затова във файла /etc/sysctl.conf се описват следните променливи на ядрото:

```
net.ipv4.ip_conntrack_max = 2000000000
net.ipv4.netfilter.ip_conntrack_max = 2000000000
```

sysctl -p

Stateless NAT

Stateless NAT (dumb NAT) е най-простата форма на NAT.

Само пренаписва адреси, преминаващи през маршрутизатора:

- входящи пакети - destination address:

```
[root@xxx-gw]# ip route add nat
205.254.211.17 via 192.168.100.17
```

- изходящи пакети - source address:

```
[root@xxx-gw]# ip rule add nat
205.254.211.17 from 192.168.100.17
```

ARP (Address Resolution Protocol)

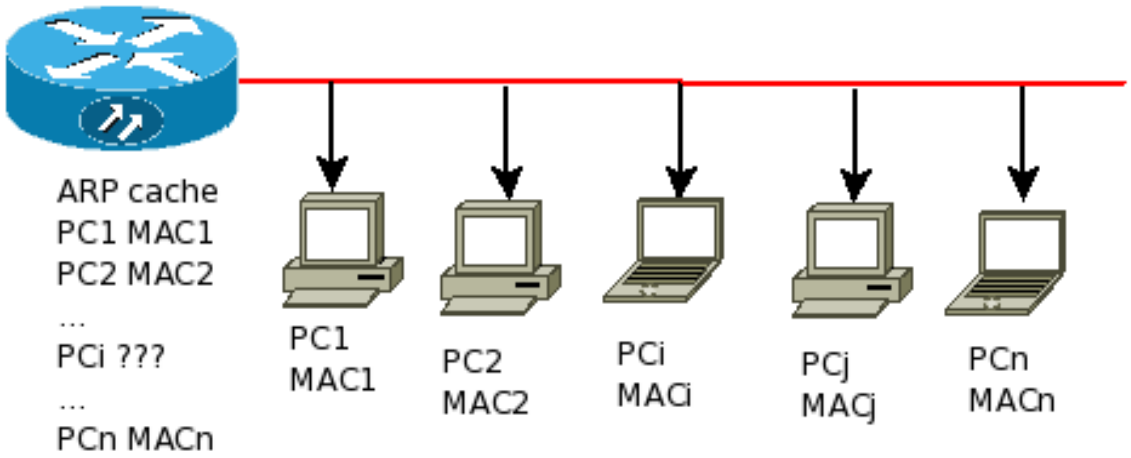
За глобална адресация в Internet се използват 32-битови IP-адреси.

В същото време хостовете, свързани към локална мрежа Ethernet, притежават уникални 48-битови MAC (физически) адреси.

При опаковането в Ethernet кадър на IP пакет, който се отправя към крайна дестинация, например, IP адресът на хоста-получател е известен,

но в полето “адрес на получателя” на Ethernet кадъра трябва да се запише Ethernet адреса на съответния хост. Иначе пакетът няма да пристигне.

ARP



За установяване на съответствието между **IP** адреса и **Ethernet** адреса на хостовете в локалната мрежа се използва протокол за право преобразуване на адресите **ARP** (address resolution protocol).

ARP cache

```
[root@shuttle ~]# arp -e
Address     Hwtype   Hwaddress     Flags Mask     Iface
loz-gw.uni-sofia.bg   ether   00:0D:56:B9:75:6D   C   eth0

[stefan@laptop ~]$ arp -e
Address   HWtype  Hwaddress   Flags Mask   Iface
192.168.0.1  ether   00:22:6b:06:d5:ad   C        wlan0
```

Complete entry - **C** flag.
Permanent entries - **M** flag.
Published entries - **P** flag.

Как работи ARP

Когато даден хост трябва да изпрати пакет (дейтаграма) към машина от локалната мрежа, чийто IP адрес е известен, но не е известен Ethernet адреса, мрежовият слой разпространява в локалната мрежа ARP пакет-заявка.

Този пакет-заявка е от тип broadcast, т.е. предава се до всички машини. В полетата “Ethernet адрес на подателя” и “IP адрес на подателя” (т.е Source IP, MAC) са записани съответните адреси на хоста, който изпраща ARP заявката.

ARP пакет

+	Bits 0 - 7	8 - 15	16 - 31
0	Hardware type (HTYPE)		Protocol type (PTYPE)
32	Hardware length (HLEN)	Protocol length (PLEN)	Operation (OPER)
64	Sender hardware address (SHA) (first 32 bits)		
96	Sender hardware address (SHA) (last 16 bits)		Sender protocol address (SPA) (first 16 bits)
128	Sender protocol address (SPA) (last 16 bits)		Target hardware address (THA) (first 16 bits)
160	Target hardware address (THA) (last 32 bits)		
192	Target protocol address (TPA)		

Как работи ARP

В полето “Данни” е записано ARP съобщение от вида “who is X.X.X.X tell Y.Y.Y.Y”, където X.X.X.X и Y.Y.Y.Y са IP адреси съответно на получателя и на подателя.

Всички машини от локалната мрежа игнорират заявката с изключение на хоста, чийто адрес съвпада с X.X.X.X.

Хост X.X.X.X изпраща ARP пакет-отговор само на подателя, тъй като вече знае неговия Ethernet адрес от получената заявка.

Как работи ARP

В полето “Данни” на пакета-отговор е записано ARP съобщение от вида “X.X.X.X is hh:hh:hh:hh:hh:hh”, където hh:hh:hh:hh:hh:hh е Ethernet адреса (в 16-ен код) на хоста, изпращащ пакета-отговор.

Обикновено хоста, който изпраща ARP заявката, запомня (кешира) получените 48-битови Ethernet адреси, за да могат да се използват при следващо предаване.

Как работи ARP

При определяне на Ethernet адреса на получателя на даден пакет първо се проверява дали този адрес не е вече кеширан

Ако не е, се изпраща ARP заявка. Хостът може да използва и адреси, записани в конфигурационен файл.

Освен това всеки хост при първоначалното си стартиране уведомява чрез broadcast съобщение от вида “I am X.X.X.X and my Ethernet address is hh:hh:hh:hh:hh:hh”, X.X.X.X и hh:hh:hh:hh:hh:hh са съответно IP адреса и Ethernet адреса.

Как работи ARP

Всички останали хостове в локалната мрежа ще запишат тази информация в своите кешове.

Чрез ARP могат да се определят физическите адреси само на хостове, които са включени в локалната мрежа и имат IP адреси от IP мрежата (подмрежата) на изпращача.

Пакетите, чийто получател е хост от друга IP мрежа (подмрежа), се изпращат към маршрутизатора, включен в локалната мрежа.

Как работи ARP

Неговият Ethernet адрес се получава чрез ARP заявка, ако не е кеширан.

Този маршрутизатор избира маршрут и препраща пакета към неговия получател.

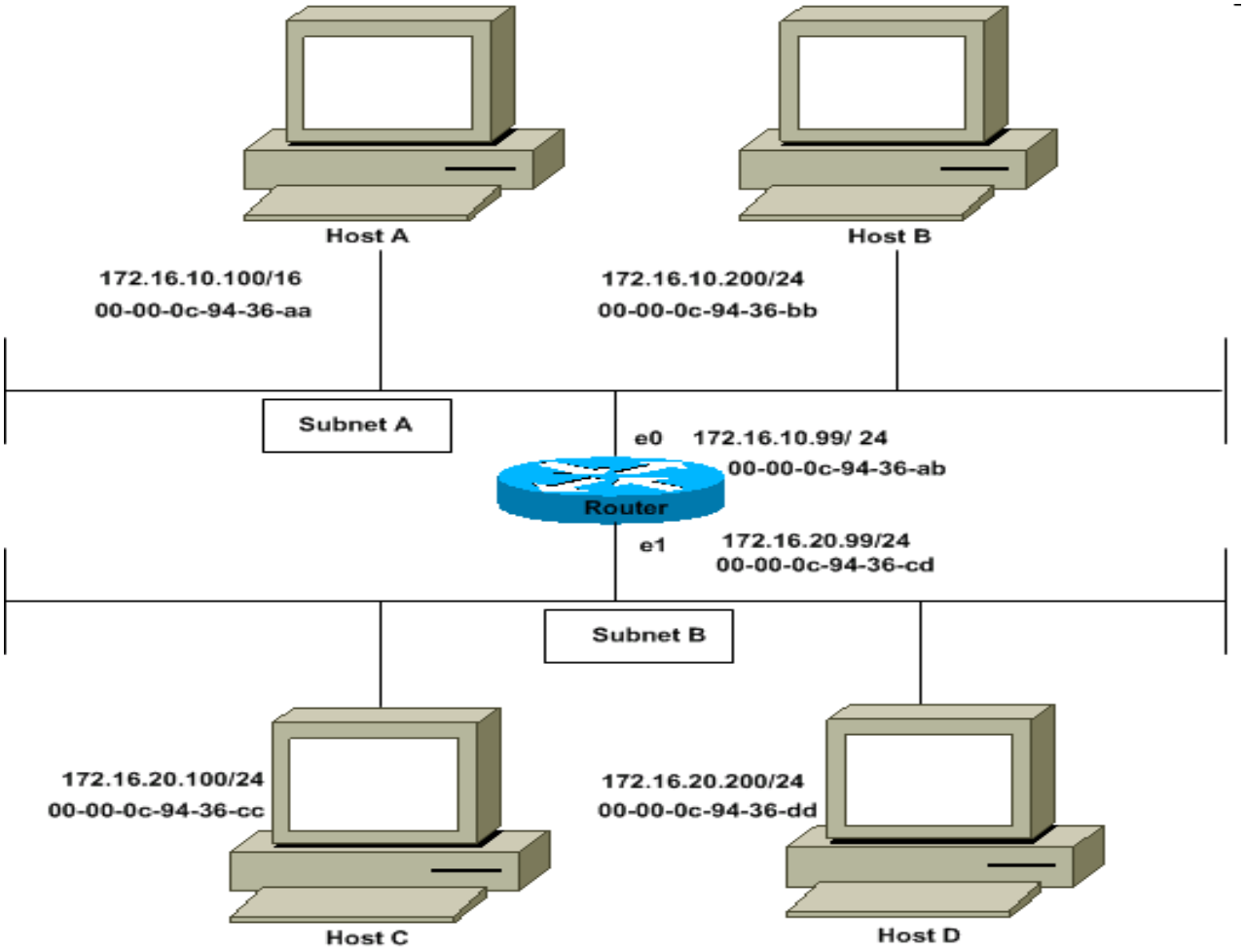
Proxy ARP

Proxy ARP е метод, чрез който хост отговаря на ARP заявки за IP адреси, които не са конфигурирани на интерфейса му.

“Проксирането” на ARP заявки за сметка на друг хост препраща целия LAN, предназначен за този хост, към прокси.

Прихванатият трафик се “превключва” към другия интерфейс на проксито (обикновено маршрутизатор) или се препраща през серийна връзка (напр., dialup или VPN тунел), за да достигне хоста получател.

Proxy ARP


Host A
Host B
172.16.10.100/16
00-00-0c-94-36-aa
172.16.10.200/24
00-00-0c-94-36-bb
Subnet A
e0 172.16.10.99/ 24
00-00-0c-94-36-ab
Router
e1 172.16.20.99/24
00-00-0c-94-36-cd
Subnet B
172.16.20.100/24
00-00-0c-94-36-cc
172.16.20.200/24
00-00-0c-94-36-dd
Host C
Host D

RARP

RARP (Reverse Address Resolution Protocol) е протокол за намиране на IP адреси по Ethernet адреси.

Обикновено IP адресът на хоста е записан в конфигурационен файл, който се намира на твърдия диск на машината.

При първоначално зареждане на операционната система файлът се прочита от твърдия диск и хостът научава своя IP адрес.

RARP

В случай, че в локалната мрежа е включена машина, която не притежава собствен твърд диск (diskless), за определяне на нейният IP адрес се използва RARP протоколът.

За целта в мрежата трябва да е включен хост, който функционира като RARP сървър.

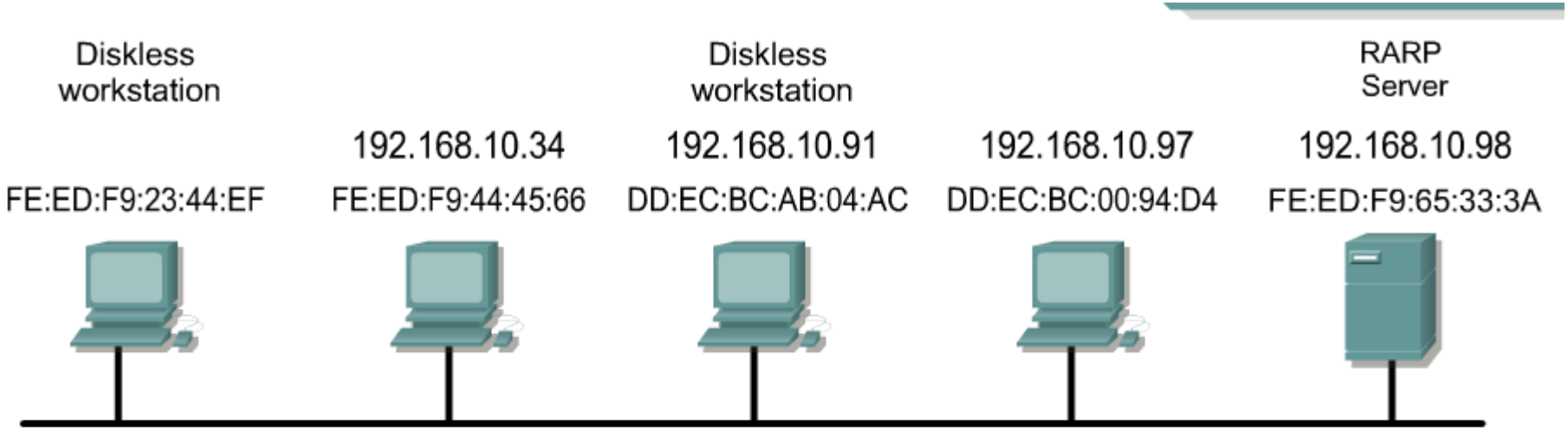
Този сървър съхранява съответствието между Ethernet и IP адреси на станциите в мрежата.

RARP

Действието на RARP се основава на наличието на уникален физически Ethernet адрес на всяка система в локалната мрежа.

При инициализиране на diskless машината RARP протоколът прочита този адрес от интерфейсната карта и предава до всички станции в мрежата пакет-заявка.

RARP сървърът отговаря на тази заявка, като в пакета-отговор се съдържа IP адреса, съответстващ на изпратения Ethernet адрес.



Frame header	2		0800_{16}
Source MAC	48	32	4
FE:ED:F9:65:33:3A	FE:ED:F9:23		
Destination MAC	44:EF		192.168
FE:ED:F9:23:44:EF	10.36		FE:ED
Field Type	F9:65:33:3A		
0X8035 (Ethernet)	192.168.10.98		

DHCP

Dynamic Host Configuration Protocol (DHCP) се използва за автоматично (динамично) конфигуриране на свързаността на даден хост към IP мрежата.

За разлика от твърдото (ръчно или статично) конфигуриране.

DHCP “раздава” не само IP адреси, но и всички други параметри на връзката – Default Gateway (изхода навън по подразбиране, DNS сървър/и, име на домейн и т.н.)

DHCP улеснява процеса на добавяне на машина в мрежата, местене и т.н.

DHCP

Днешната версия на DHCP за IPv4 е стандартизирана в RFC 2131 (1997 г.).

DHCP за IPv6 (DHCPv6) е дефинирана в RFC 3315.

DHCP е протокол от типа клиент-сървър.

DHCP-конфигуриран клиент веднага след включването се свързва към мрежата и изпраща broadcast заявка, искайки необходимата информация от DHCP сървър.

DHCP

DHCP сървърът разполага с пул от IP адреси и необходимата информация за конфигуриране на клиента: GW, SM, домейн, DNS сървър/и, NTP, WINS и др.

При получаване на валидна заявка сървърът присвоява IP адрес, време за отдаване на адреса (lease time – през което алокацията е валидна) и др. (гореспоменати) IP конфиг. параметри.

Раздаване на IP адреси (allocation)

DHCP сървърите раздават (алокират) IP адреси по 3 начина:

Динамична алокация: Обхват от IP адреси се дават за DHCP и всеки клиент си заявява IP адрес от DHCP сървъра при включване. Времето на отдаване (lease) е дефинирано, така че сървърът може да преотдаде адреса на друга машина.

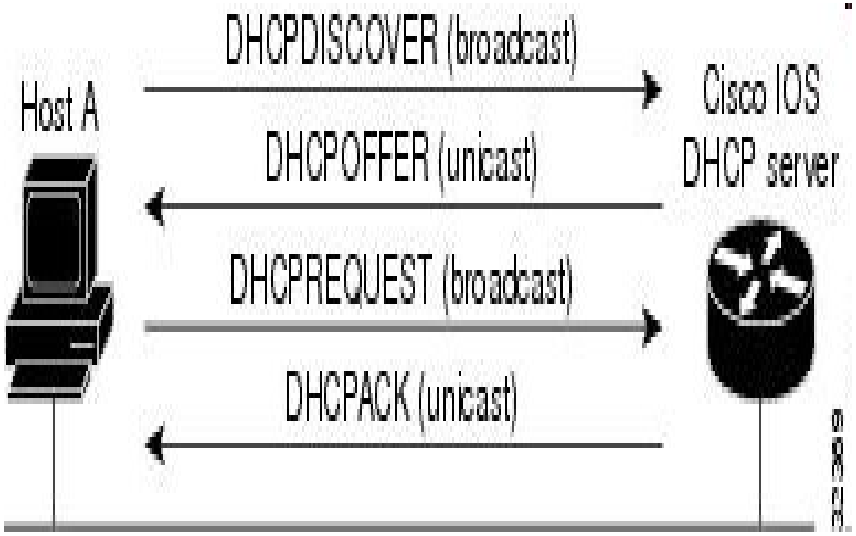
Автоматична алокация: Подобн е на динамичната, но даден IP адрес е резервиран за даден клиент.

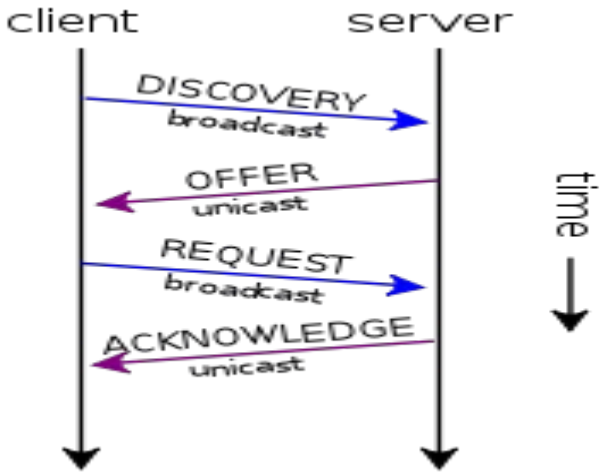
Раздаване на IP адреси (allocation)

Статична алокация: DHCP раздава IP адреси на базата на таблица MAC адрес/IP адрес, ръчно попълнена от администратора. Само клиенти, чиито MAC адреси присъстват в тази таблица, ще получат IP адреси.

Нарича се още *Static DHCP Assignment* (от DD-WRT, Linux-базиран фърмуер, Linksys), *fixed-address* (от dhcpd), *DHCP reservation* или *Static DHCP* (от Cisco/Linksys) или *IP reservation*, *MAC/IP binding* (други производитевли).

Фази на DHCP процеса





DHCP discovery

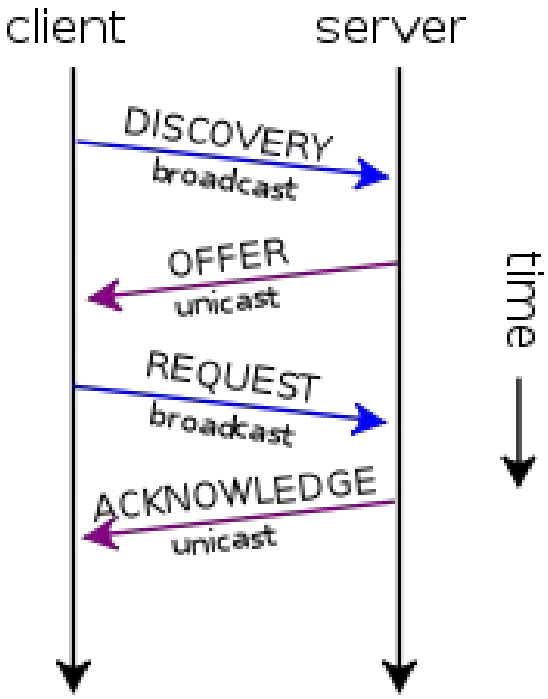
DHCP сървър:
62.44.109.140

DHCP пул: 62.44.109.141 – 254/25

Последен свободен адрес:
62.44.109.151

Транзакция 654:

UDP Src=0.0.0.0 sPort=68
Dest=255.255.255.255
dPort=67



DHCP offer

UDP Src=62.44.109.140 sPort=67
Dest=255.255.255.255 dPort=68

Offer IP: 62.44.109.151

ID: 654

Lease Time: 3600 s

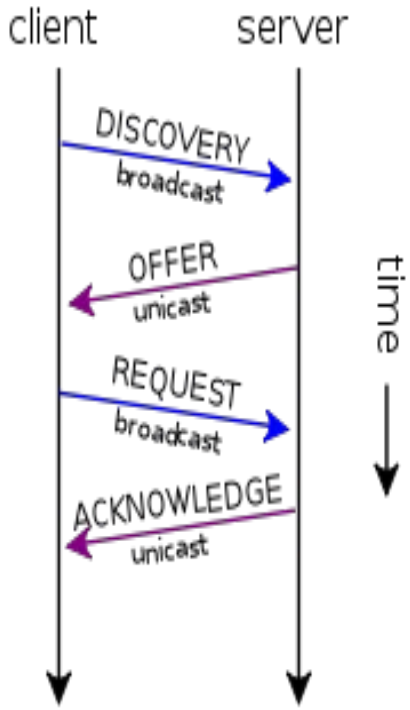
SM: 255.255.255.0

DHCP server: 62.44.109.140

Router (GW): 62.44.109.193

DNS: 62.44.109.1, 62.44.96.1

Domain: ucc.uni-sofia.bg



DHCP request

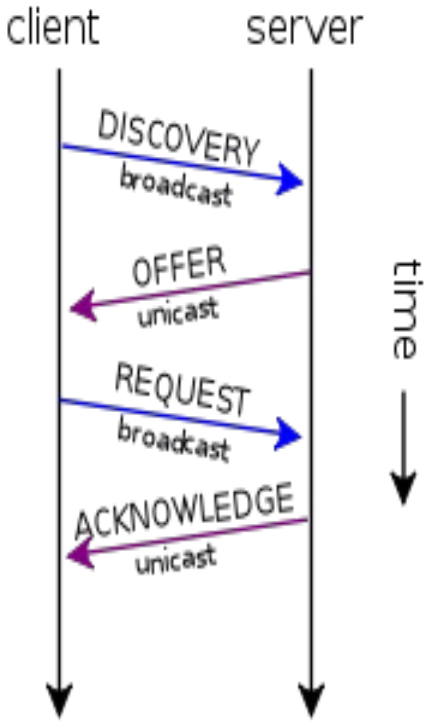
UDP Src=0.0.0.0 sPort=68
Dest=255.255.255.255
dPort=67

Requested IP:
62.44.109.151

ID: 655

DHCP server:
62.44.109.140

Lease Time: 3600 s



DHCP acknowledgement

UDP Src= 62.44.109.140 67
Dest=255.255.255.255 68

Requested IP: 62.44.109.151

ID: 655

Lease Time: 3600 s

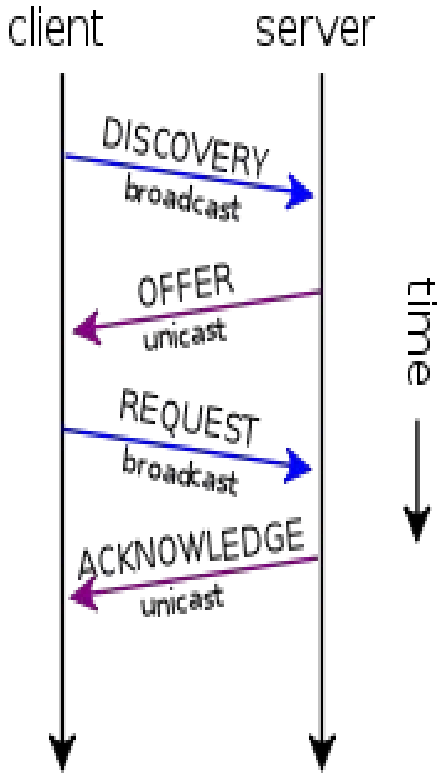
SM: 255.255.255.0

DHCP server: 62.44.109.140

Router (GW): 62.44.109.193

DNS: 62.44.109.1, 62.44.96.1

Domain: ucc.uni-sofia.bg

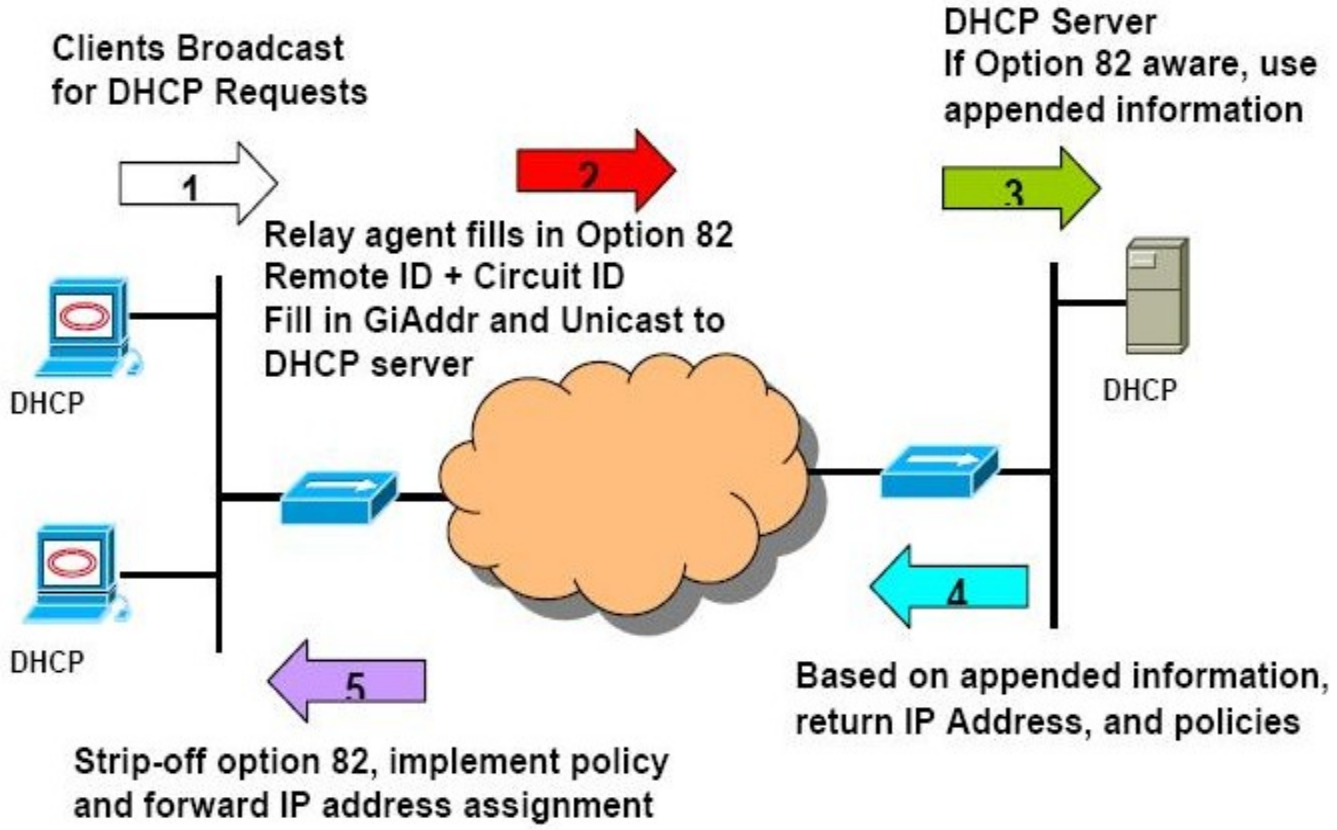


DHCP Relay

Желателно е DHCP сървър и клиенти да са на един и същ сегмент (Ethernet и IP).

Когато това не е възможно, прилага се DHCP Relay.

DHCP Relay


Clients Broadcast
for DHCP Requests
1
2
Relay agent fills in Option 82
Remote ID + Circuit ID
Fill in GiAddr and Unicast to
DHCP server
DHCP Server
If Option 82 aware, use
appended information
3
DHCP
DHCP
DHCP
4
Based on appended information,
return IP Address, and policies
5
Strip-off option 82, implement policy
and forward IP address assignment

vim /etc/dhcpd.conf

```
subnet 172.18.0.0 netmask 255.255.254.0 {

# --- default gateway
        option routers                  172.18.0.1;
        option subnet-mask              255.255.254.0;

        option nis-domain               "uni-sofia.bg";
        option domain-name              "conf.uni-sofia.bg";
        option domain-name-servers
    62.44.96.7,62.44.96.1;
```

vim /etc/dhcpd.conf (cont'd)

```
    option time-offset              7200;   # East European
  Standard Time
    option ntp-servers              62.44.96.44;
#    option ntp-servers              62.44.96.7, 62.44.96.1;
#    option netbios-name-servers     192.168.1.1;
# --- Selects point-to-point node (default is hybrid). Don't
  change this unless you understand Netbios very well
#    option netbios-node-type 2;
```

vim /etc/dhcpd.conf (cont'd)

```
host vlado {
   option host-name "vladi";
   hardware ethernet 00:0a:e4:b1:6e:52;
   fixed-address 172.18.0.101;
}

range 172.18.0.2 172.18.1.254;
}
```

ICMP

Internet Control Message Protocol (**ICMP**) е част от протокола IP.

Използва се от мрежовите ОС главно за откриване на грешки по мрежата и изпращане на съобщения за това.

ICMP за IPv4 е известени като ICMPv4. IPv6 има подобен, ICMPv6.

Дефиниран е в RFC 792.

IP опакова ICMP съобщението с нов IP хедър, за да го върне на изпращача и го предава като обикновен пакет.

ICMP

Например, всеки възел в мрежата (рутер, GW), която направлява IP пакета, трябва да декрементира TTL полето на IP хедъра с 1.

Ако TTL достигне 0, ICMP съобщение Time to live exceeded in transit message се изпраща към източника.

ICMP съобщенията се съдържат в стандартни IP пакети, но се обработват като специални случаи.

Много мрежови средства за диагностика се базират на ICMP.

ICMP

Командата traceroute изпраща UDP дейтаграми с определени IP TTL полета и очаква ICMP Time to live exceeded in transit, също изпраща "Destination unreachable" в отговор.

Средството ping се реализира с ICMP "Echo request" и "Echo reply" съобщения.

Структура на ICMP пакет

	Bit 0 - 7	Bit 8 - 15	Bit 16 - 23	Bit 24 - 31
IP Header (160 bits OR 20 Bytes)	Version/IHL	Type of service	Length	
	Identification		*flags* and *offset*	
	Time To Live(TTL)	Protocol	Checksum	
	Source IP address			
	Destination IP address			
ICMP Payload (64+ bits OR 8+ Bytes)	Type of message	Code	Checksum	
	Quench			
	Data (*optional*)			

Структура на ICMP пакет

Тип – вж. по-долу.

Код - вж. по-долу.

Checksum – контролна сума за ICMP header+data.

Данни

Linux ping 56 байта (октета) плюс 8 за хедър.

Windows "ping.exe" - 32 + 8 хедър.

ICMP съобщения

Type	Code	Description
0 - Echo Reply	0	Echo reply (used to ping)
1 and 2		*Reserved*
3 - Destination Unreachable	0	Destination network unreachable
	1	Destination host unreachable
	2	Destination protocol unreachable
	3	Destination port unreachable
	4	Fragmentation required, and DF flag set
	5	Source route failed
	6	Destination network unknown
	7	Destination host unknown
	8	Source host isolated
	9	Network administratively prohibited
	10	Host administratively prohibited
	11	Network unreachable for TOS
	12	Host unreachable for TOS
	13	Communication administratively prohibited
4 - Source Quench	0	Source quench (congestion control)

ICMP съобщения

5 - Redirect Message	0	Redirect Datagram for the Network
	1	Redirect Datagram for the Host
	2	Redirect Datagram for the TOS & network
	3	Redirect Datagram for the TOS & host
6		Alternate Host Address
7		*Reserved*
8 - Echo Request	0	Echo request
9 - Router Advertisement	0	Router Advertisement
10 - Router Solicitation	0	Router discovery/selection/solicitation
11 - Time Exceeded	0	TTL expired in transit
	1	Fragment reassembly time exceeded
12 - Parameter Problem: Bad IP header	0	Pointer indicates the error
	1	Missing a required option
	2	Bad length
13 - Timestamp	0	Timestamp
14 - Timestamp Reply	0	Timestamp reply
15 - Information Request	0	Information Request
16 - Information Reply	0	Information Reply
17 - Address Mask Request	0	Address Mask Request
18 - Address Mask Reply	0	Address Mask Reply
19		*Reserved* for security
20 through 29		*Reserved* for robustness experiment
30 - Traceroute	0	Information Request

ping

Ping е инструмент за тестване на достижимостта на даден хост по IP мрежата.

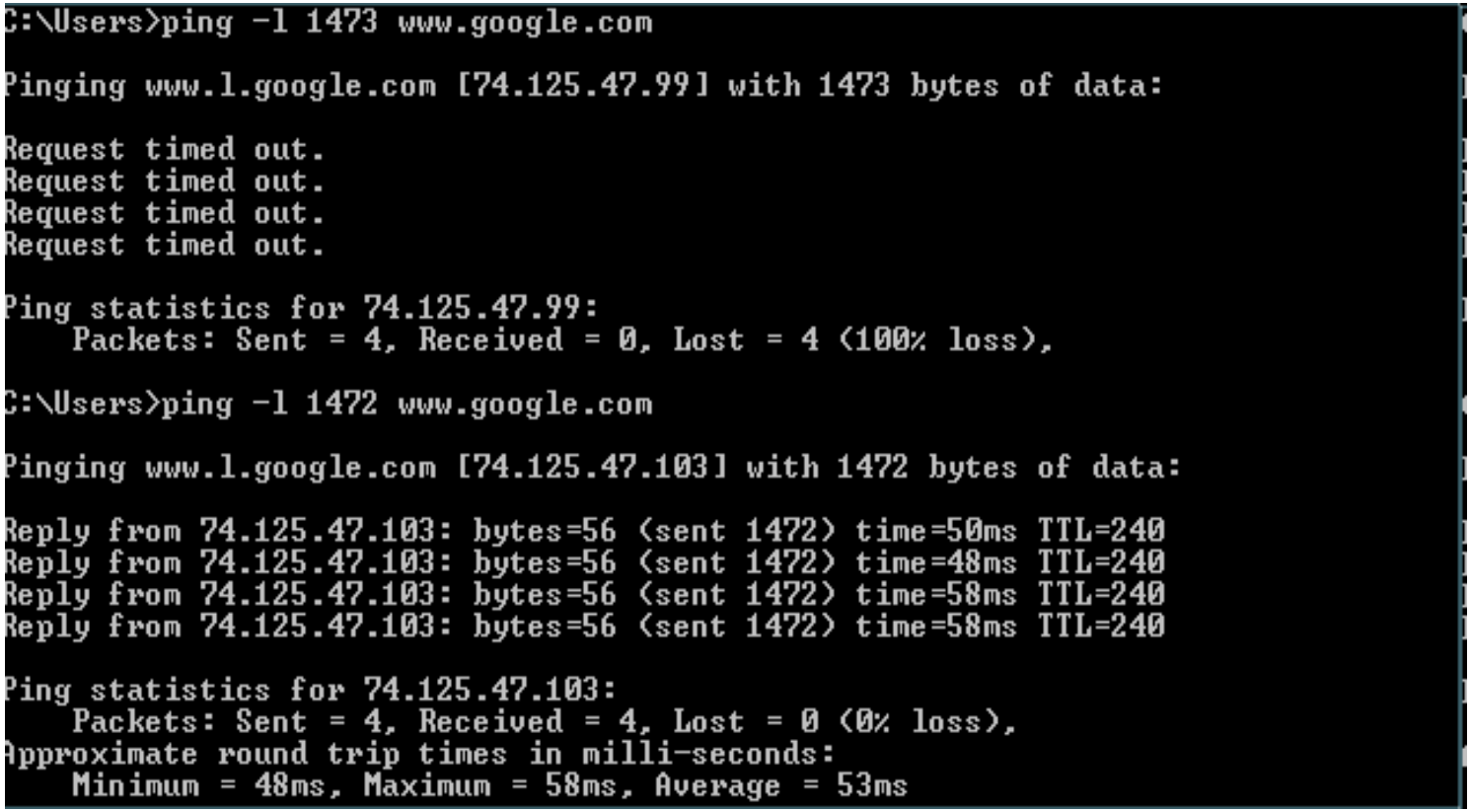
Изпраща ICMP “echo request” пакети към целта и очаква ICMP “echo response” отговори.

Ping измерва round-trip time и регистрира загуби на пакети.

Накрая разпечатва статистика: минималното, средното, максималното и (в някои версии) стандартното отклонение от round trip time.

Mike Muuss е написал програмата през декември, 1983. Нарекъл я на звуковите импулси, издавани от локатор в подводница.

Пример на ping



traceroute

traceroute е инструмент за определяне на маршрута на пакетите по мрежата. За IPv6 вариантът е **traceroute6**.

traceroute го има за всички Unix-подобни ОС. Подобна функционалност имат **tracepath** на модерните Linux дистрибуции и **tracert** в Microsoft Windows.

Traceroute инкрементира с 1 "time-to-live" (TTL) на всяка следваща “тройка” от изпратени пакети. Първата тройка е с TTL=1. Следващата е с TTL = 2 и т.н. Минавайки през хост, TTL на пакета се декрементира с 1 и се отправя към следващия хост. Хостът изхвърля пристигнал пакет с TTL = 1 и изпраща на подателя ICMP t ime exceeded (type 11).

traceroute използва тези връщани пакети, за да създаде списък от хостове, през които пакетът е минал по маршрута до дестинацията.

traceroute

Трите timestamp за всеки хост по пътя са закъснението - delay (latency) в ms за всеки пакет от тройката.

Ако пакетът не се върне в рамките на очаквания timeout, разпечатва се звездичка (asterisk).

Traceroute може и да не изброи реалните хостове. Само показва, че първият хост е на един хоп, вторият – на два, и т.н.

Просто IP не гарантира, че всички пакети ще минат по един и същ път.

Ако хост на хоп N не отговори, този хоп ще бъде пропуснат в разпечатката.

traceroute

В съвременните Unix и Linux-базирани ОС traceroute използва по подразбиране UDP дейтаграми с номера на дестинационни портове 33434 - 33534. Но има опция да се използва ICMP echo request (type 8) както в Windows tracert.

traceroute

```
[root@shuttle ~]# traceroute ripe.net
traceroute to ripe.net (193.0.19.25), 30 hops max, 40 byte packets
 1  ucc-gw.ucc.uni-sofia.bg (62.44.109.5)  0.227 ms  0.237 ms  0.252 ms
 2  border-main.uni-sofia.bg (62.44.127.21)  0.590 ms  0.584 ms  0.567 ms
 3  core-su.lines.acad.bg (194.141.252.21)  1.179 ms  1.311 ms  1.448 ms
 4  istf.rt1.sof.bg.geant2.net (62.40.125.141)  1.266 ms  1.265 ms  1.232 ms
 5  so-2-3-0.rt1.bud.hu.geant2.net (62.40.112.202)  15.477 ms  15.490 ms  15.468
 ms
 6  bpt-b2-link.telia.net (80.239.134.1)  15.437 ms  14.885 ms  14.944 ms
 7  hbg-bb1-link.telia.net (80.91.250.130)  36.623 ms  36.597 ms  36.586 ms
 8  adm-bb1-link.telia.net (80.91.252.40)  46.064 ms adm-bb1-link.telia.net (80.
91.253.45)  44.637 ms  44.634 ms
 9  adm-b1-link.telia.net (80.91.254.221)  44.634 ms adm-b2-link.telia.net (80.9
1.254.133)  44.818 ms  44.821 ms
10  * gw.amsix.nikrtr.ripe.net (195.69.144.68)  468.713 ms *
11  gw.transit.nsrp.ripe.net (193.0.3.1)  40.922 ms  42.555 ms  40.804 ms
12  aquila.ripe.net (193.0.19.25)  43.185 ms  41.560 ms  43.088 ms
```

arping

arping е подобна на ping, но използва ARP вместо ICMP.

Затова, arping е използваема само в локалната мрежа

В някои случаи отговорът може да идва от междинна система - proxy ARP (напр. рутер).

arping

```
[root@shuttle ~]# arping 62.44.109.1
ARPING 62.44.109.1 from 62.44.109.11 eth0
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.638ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.608ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.604ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.610ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.594ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.591ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.585ms
Unicast reply from 62.44.109.1 [00:40:95:30:13:ED]  0.591ms
Sent 8 probes (1 broadcast(s))
Received 8 response(s)
```