NETWORKERS

CISCO SYSTEMS

Understanding DNS and DHCP

Session NSC-141

Agenda

- **Addresses**
  - Hierarchy and Topology
  - Assignment and Reliability
- **Naming**
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

Agenda

Cisco.com

- **Addresses**
  - **Hierarchy and Topology**
  - **Assignment and Reliability**
- **Naming**
  - **Structure**
  - **Protocol**
  - **Reliable Operation**
  - **New Things**

Address Review

Cisco.com

128	64	32	16	8	4	2	1	128	64	32	16	8	4	2	1	128	64	32	16	8	4	2	1	128	64	32	16	8	4	2	1
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

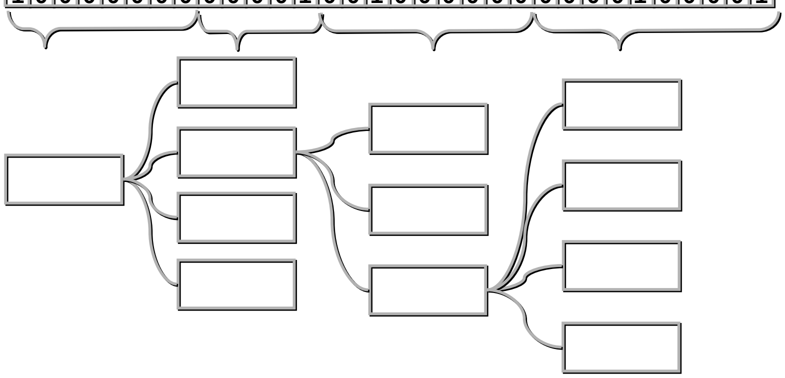
128	9	0	33

- **IPv4 address 32 bits (e.g. 128.9.0.33)**
- **Dotted decimal~8-bit fields**
- **IPv6 128 bits~colon-separated hexadecimal 16-bit fields**

Address Hierarchy

128.9.0.33

1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1



- **Best if topology matches hierarchy**

Subnet Mask

Address 128.9.0.33

1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

Mask 255.255.255.0

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0

- **Mask separates network (1) from host (0) part of the address**
- **Prefix (longest match) routing—contiguous “1” bits to the left**

Subnets

Cisco.com

- **Each range of addresses for hosts defines a subnet e.g. 128.9.0.0/24**
- **Within the subnet, hosts communicate directly, using layer 2**
- **Special meaning for certain addresses**

  All-ones—broadcast

  All-zero—network

Special Addresses

Cisco.com

- **Multicast**
  IPv4—224-239.d.d.d [RFC 2365]
  IPv6—FFxx:x:x:x:x:x:x:x
- **Any-cast [RFC 1546]**
  looks like unicast, but with multiple advertisers
- **Site-local**
  IPv4—10/8, 172.16/12, 192.168/16 [RFC 1918]
  IPv6—FEC0:0:0:<subnet ID>:<interface ID>
- **Link-local**
  IPv4: 169.254/16 IPv6: FE80:0:0:0:<interface ID>
- **Loop-back—127.0.0.1–0:0:0:0:0:0:0:1 (::1)**

Agenda

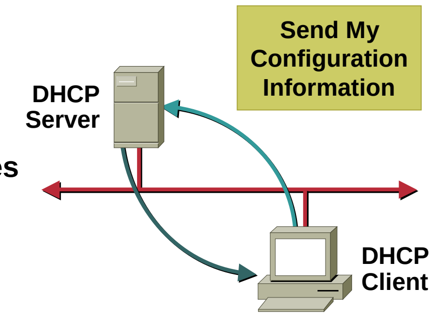
- **Addresses**
  - Hierarchy and Topology
  - Assignment and Reliability
- **Naming**
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

DHCP Bootstrapping

- Ideal administrator
- Assignment is temporary
- Backup for reliability
- Automation avoids personality

How DHCP Works Obtaining a Lease

- Dynamically assigns configuration information
- Creates IP address pools to conserve addresses and support mobile users
- Clients broadcasts DHCP DISCOVER packet on local subnet
- Multiple servers can respond
- Client chooses first or best response

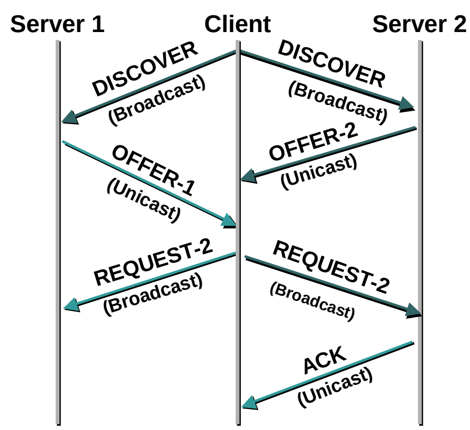


Here is Your Configuration:
IP Address: 192.204.18.7
Subnet Mask: 255.255.255.0
Default Routers: 192.204.18.1, 192.204.18.3
DNS Servers: 192.204.18.8, 192.204.18.9
Lease Time: 5 days

How DHCP Works

DHCP Discover Process

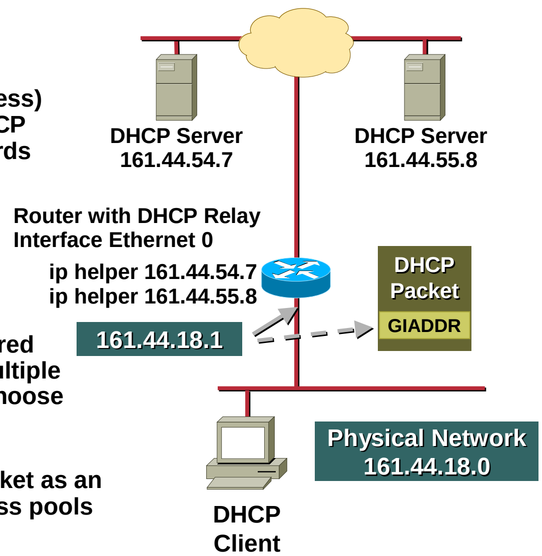
- DHCP client broadcasts DHCP DISCOVER packet on local subnet
- DHCP servers send OFFER packet with lease information
- DHCP client selects lease and broadcasts DHCP REQUEST packet
- Selected DHCP server sends DHCP ACK packet



DHCP Relay

Cisco.com

- DHCP clients broadcasts a DHCP discover packet
- DHCP relay (ip helper address) on the router hears the DHCP Discover packet and forwards (unicast) the packet to the DHCP server
- DHCP relay fills in the GIADDR field with IP address of the primary interface of router
- DHCP relay can be configured to forward the packet to multiple DHCP servers; client will choose the "best" server
- DHCP servers use GIADDR field of DHCP Discover packet as an index in to the list of address pools

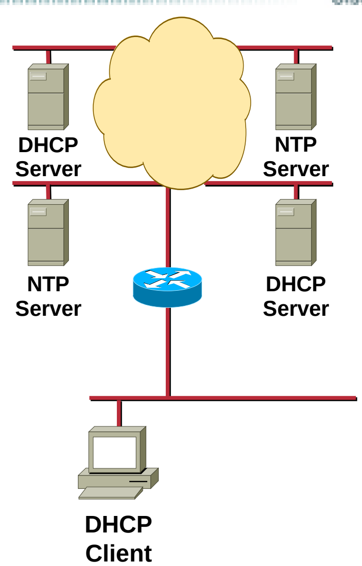


NSC-141
5043_04_2002_c2

DHCP Options for Applications

Cisco.com

- Server location [RFC 2610]
- Novell directory services [RFC 2241]
- Time, NIS, TCP and IP parameters… [RFC 2131]



NSC-141
5043_04_2002_c2

NSC-141, 5043_04_2002

DHCP Reliability

Cisco.com

- **Failover**

  Draft based on our design

  Coordinating the name in the DNS uses an identifier stored in DNS

- **Loadsharing**

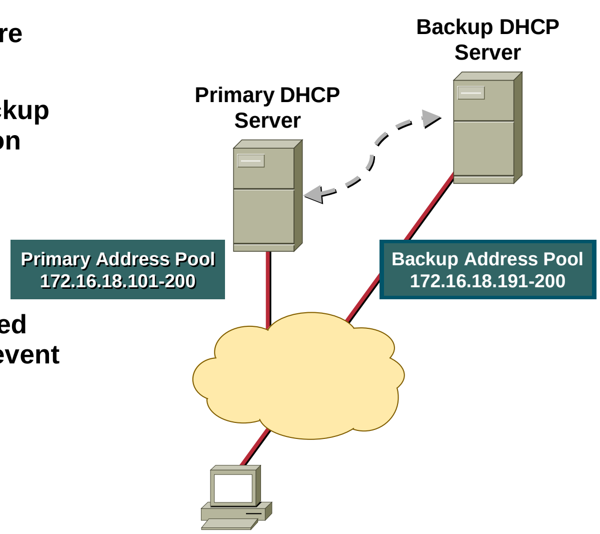
  Servers answer only for configured hash (MAC)

  RFC 3074

DHCP Safe Failover Protocol

Cisco.com

- **All DHCP requests are sent to both servers**
- **Primary updates backup with lease information**
- **Backup takes over when primary fails**
- **Backup server uses dedicated pool of addresses allocated by the primary to prevent duplicate IP address**
- **Servers synchronize when primary is up**
- **IETF Internet draft**



How DHCP Works DHCP Packet

Cisco.com

OP Code	Hardware Type	Hardware Length	HOPS
Transaction ID (XID)			
Seconds		Flags	
Client IP Address (CIADDR)			
Your IP Address (YIADDR)			
Server IP Address (SIADDR)			
Gateway IP Address (GIADDR)			
Client Hardware Address (CHADDR)—16 bytes			
Server Name (SNAME)—64 bytes			
Filename—128 bytes			
DHCP Options			

Summary

Cisco.com

- DHCP
- Questions?

Agenda

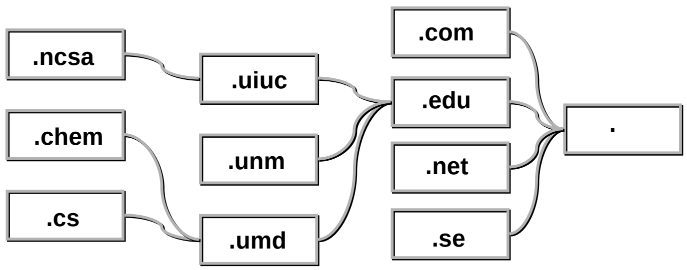
- Addresses
  - Hierarchy and Topology
  - Assignment and Reliability
- Naming
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

Domain Name Service

- DNS is a database
  - And the protocol to access it
- Distinctive features:
  - Design for look-up queries
  - Replicated content
  - Distributed control (zones)

Name Hierarchy

- Independent of address hierarchy
- Names length not limited by address size (63 bytes/label, 255 bytes/FQDN)



Terminology

- Label (name, owner)
- Resource record (type)
- Value (encoded by type)

Record Format

Label **RR-Type** **Value**

<name>	[<ttl>]	[<class>]	<type>	<data>
VAXA.ISI.EDU.		IN	A	10.2.0.27
VAXA.ISI.EDU.		IN	A	128.9.0.33

Optional Fields:
We Only Care About Class=IN (Internet)
ttl Will Get Attention Later

Address Examples

Label	RR-Type	Value
<name>	<type>	<data>
VAXA.ISI.EDU.	A	10.2.0.27
	A	128.9.0.33

Empty Label Is the Same as Line Above

Address and Canonical Name

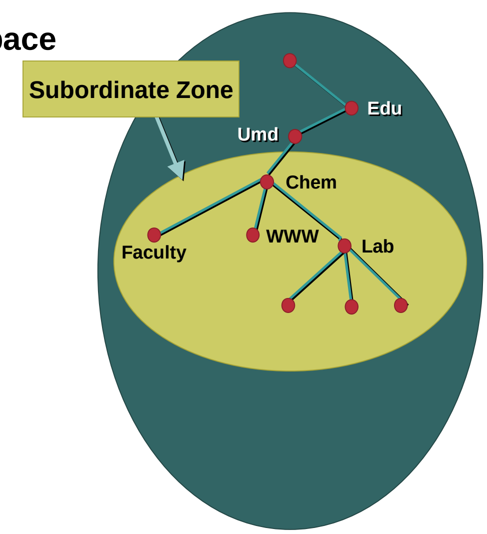
- 'A' resource record (RR)
- 'CNAME' RR

  The value of a canonical name is ALWAYS the label of an A record

- Service alias using CNAME [RFC 2219] Best Current Practice

Delegation Zone

- Hierarchical name space
- Each node in tree represents domain/subdomain
- Some subdomains are defined as zones
- Each zone has a "primary"name server responsible for all lower nodes
- Resource Records (RR) are defined for each node



Delegation Records

- Distributes database administration
- Name Server (NS) RR
- Zone Start of Authority (SOA)

Delegation: NS and "Glue"

- NS Resource Record (RR)
- "Glue" entries in parent and child

<domain> NS <server>

```
SRI.COM.      NS  KL.SRI.COM.
KL.SRI.COM.   A   10.1.0.2
```

Delegation: SOA

Cisco.com

```
<name> [<ttl>] [<class>] SOA <origin> <person>
  ( <serial>
    <refresh>
    <retry>
    <expire>
    <minimum> )
```

```
$ORIGIN ARPA.

@ IN SOA  SRI-NIC.ARPA.  HOSTMASTER.SRI-NIC.ARPA. (
              45          ;serial (sequential)
              3600        ;refresh (1 hour regular check)
              600         ;retry (10 minutes until check)
              3600000     ;expire (42 days until refresh)
              86400 )     ;minimum (a day)
```


Reverse

Cisco.com

- **Another hierarchy for in-addr.ARPA**

  Reverse the order in the label because names aggregate within suffixes rather than (address) prefixes

```
27.0.2.10.IN-ADDR.ARPA.    PTR   VAXA.ISI.EDU.
33.0.9.128.IN-ADDR.ARPA.   PTR   VAXA.ISI.EDU.
```

Reverse Requirement

Cisco.com

> **"For every IP address, there should be a matching PTR record in the in-addr.arpa domain."**
>
> **RFC 1912,**
> **Common DNS Operational and Configuration Errors**

NSC-141
5043_04_2002_c2

Reverse Not CNAME

Cisco.com

> **"PTRs should use official names and not aliases."**
>
> **RFC 1033**
> **Domain Administrators Operations Guide**

NSC-141
5043_04_2002_c2

Reverse Complication

Cisco.com

- Address field separations not on dotted decimal boundaries
- Create CNAMEs in IN-ADDR.ARPA
  - Labels are just text—describing addresses
  - Parent creates labels: 1.2/25…in-addr.arpa
  - Delegates zone for 2/25…in-addr.arpa

NSC-141
5043_04_2002_c2 35

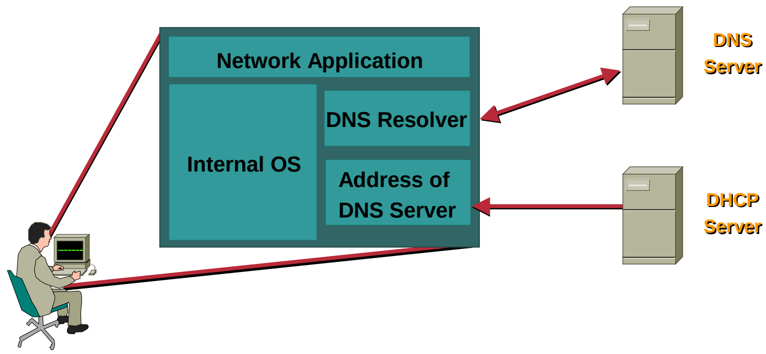
Agenda

Cisco.com

- Addresses
  - Hierarchy and Topology
  - Assignment and Reliability
- Naming
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

NSC-141
5043_04_2002_c2 36

DNS Servers and Resolvers



- Application connects by name, the OS gets the address from the resolver
- Most applications use addresses in the order provided by the resolver

TCP and UDP Ports

- Port 53 for both TCP and UDP
- UDP for queries if small enough
- TCP for zone transfer
- Server can use source port of 53 when “forwarding”

Redirection and Recursion

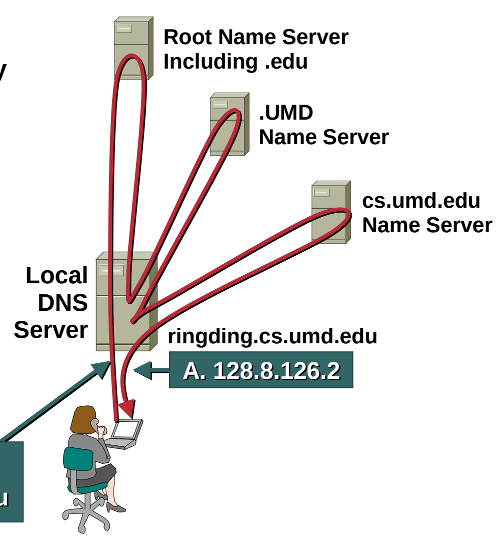
Cisco.com

- **Redirection:**
  “Take your question down the hall”
- **Recursion:**
  “I’ll get back to you”
- **Resolver sets recursion desired (RD), server responds with recursion available (RA) through bits in the DNS header**

DNS First Query

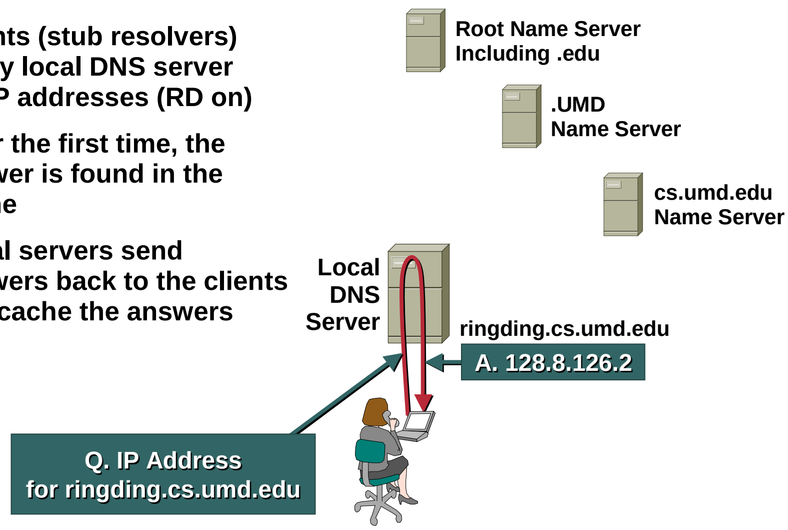
Cisco.com

- **Clients (stub resolvers) query local DNS server for IP addresses (RD on)**
- **Local server queries (RD off) the root name server and follows referrals until it finds a server that has the answer**
- **Local servers send answers back to the clients and cache the answers**



DNS Subsequent Queries

- Clients (stub resolvers) query local DNS server for IP addresses (RD on)
- After the first time, the answer is found in the cache
- Local servers send answers back to the clients and cache the answers



Caching and Forwarders

- Caching is controlled by the Time to Live (TTL) in RR
- Negative caching required by RFC 2308
- The “minimum” TTL parameter in the SOA (with the ttl of the SOA RR itself) determines the TTL for caching negative answers
- Sending a recursive query to a forwarder builds a cache for the site

Time to Live

Cisco.com

- Changing host addresses
- Reduce TTL prior to change
- Then restore to manage the load
- CNR dynamically updates DNS TTL with 1/3 DHCP lease time

NSC-141
5043_04_2002_c2

Agenda

Cisco.com

- Addresses
  - Hierarchy and Topology
  - Assignment and Reliability
- Naming
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

NSC-141
5043_04_2002_c2

Secondary Servers

- Reliability depends on separation
- Location—physical and subnet
- Independent fate—separate power
- Separate administration if possible
- RFC 2182—best current practice

Replication

- Transfer zone contents (AXFR)
- Transfer controlled by serial number and refresh parameters in SOA

Replication Efficiency

Cisco.com

- Notify (new protocol operation) enables primary to inform secondary when RRs have changed [RFC 1996]
- Incremental transfer (IXFR) sends just changes to the zone [RFC 1995]

Agenda

Cisco.com

- Addresses
  - Hierarchy and Topology
  - Assignment and Reliability
- Naming
  - Structure
  - Protocol
  - Reliable Operation
  - New Things

Dynamic Update

- Atomic update of RR-set
- Only if conditions are met
  - RR exists
  - RR has specified value
- Trusting version—RFC 2136
- Secure version—RFC 3007
- Created so that DHCP can update DNS

Securing Queries

- TSIG—Transaction Signature
  RFC 2845
- Secret-key hash of the transaction (HMAC-MD5) to the forwarder
- Pseudo RR, not cached or saved
- Only useful with local forwarders

Securing Zone Transfer

- TSIG RR
- Secondary servers have an administrative relationship that can support secret keys
- Don’t need the overhead of public keys

TKEY

- Transaction KEY—not stored
- Use DNS to establish secret keys alternative to manual keys
- Modes include

  Diffie-Hellman

  GSS-API

  Server or Resolver assigned encrypted (encrypted using KEY RR)
- RFC 2930

SIG (0)

Cisco.com

- Use DNS for client to authenticate to server
- Authenticates the transaction
- Public KEY in DNS
- Private key in client
- RFC 2931

Securing Zone Contents

Cisco.com

- DNS security
- Key RR—distributes public keys for records
- SIG RR—authenticates (signs) one RR set
- NXT RR—“next” record enables authentication of non-existence
- DS RR—delegation signer~key in parent with which the child zone is signed
- RFC 2535—being revised

Deployment of DNS Sec

Cisco.com

- Experimental only now
- Trust depends on the entire path from the resolver
- Signing all the RRsets in .com is an unresolved problem

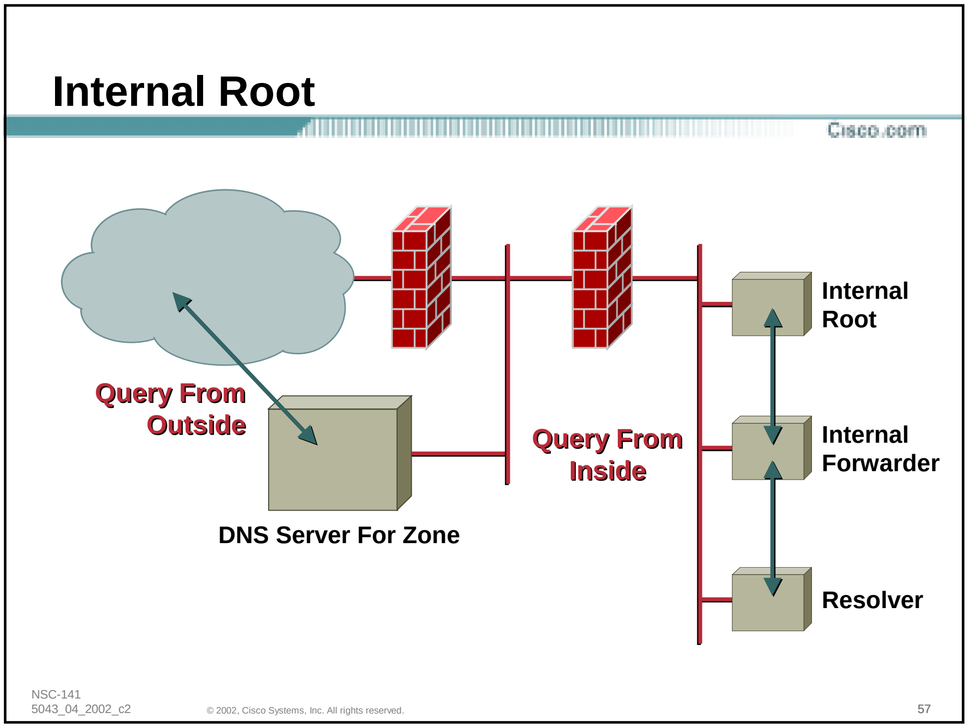
Split DNS

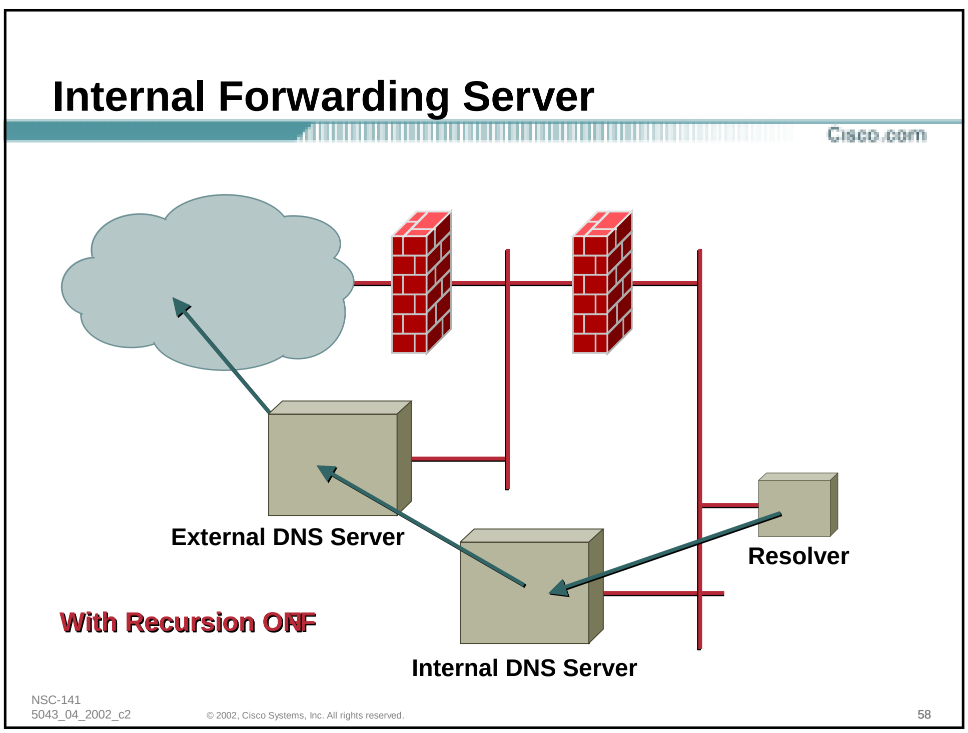
Cisco.com

- External holds limited contents for public
- Internal

  Isolated clients query DNS servers configured as root

  Internal (secondary) servers forward to external caching server for other domains


Internal Root
Cisco.com
Internal
Root
Query From
Outside
Query From
Inside
Internal
Forwarder
DNS Server For Zone
Resolver
NSC-141
5043_04_2002_c2
© 2002, Cisco Systems, Inc. All rights reserved.
57


Internal Forwarding Server
Cisco.com
External DNS Server
Resolver
With Recursion ONF
Internal DNS Server
NSC-141
5043_04_2002_c2
© 2002, Cisco Systems, Inc. All rights reserved.
58

IP Version 6

Cisco.com

- **AAAA—RFC 3152—IP6.ARPA**

  ARPA~Addressing and Routing Parameters Area

```
v6host.example.com. AAAA 4321:0:1:2:3:4:567:89ab

b.a.9.8.7.6.5.0.4.0.0.0.3.0.0.0.2
  .0.0.0.1.0.0.0.0.0.0.0.0.1.2.3.4.IP6.ARPA.
  PTR v6host.example.com.
```


Records for Applications

Cisco.com

- **MX**
- **SRV**
- **NAPTR**

MX

Cisco.com

- Mail eXchange RR
- Where the mail for the host is to be sent
- Round-robin within equal preferences
- Mailers send only to lower preference numbers

Name	TTL	Class	MX	Preference	Target
BAZ.FOO.COM.			MX	10	PO1.FOO.COM.
			MX	20	PO2.FOO.COM.
			MX	30	PO3.FOO.COM.

Wildcards

Cisco.com

- Special treatment for ‘*’ in the label
- Any name in the query matches, and the answer is synthesized
- Most often used in mail exchange

```
FOO.COM.       MX   10    RELAY.CS.NET.
*.FOO.COM.     MX   20    RELAY.CS.NET.
```

SRV

- Generalize the MX idea
- Find hosts offering service in a domain
- Add structure to the name
- Add fields to the RR—specialize priority and weight (replace preference)
- Target must not be an alias (CNAME)
- RFC 2782

NSC-141
5043_04_2002_c2

SRV

Format of RR and Example

```
_Service._Proto.Name    SRV Priority Weight Port Target
_ldap._tcp.example.com    SRV  1  10  389  ldap1.example.com
                          SRV  1  20  389  ldap2.example.com
```

NSC-141
5043_04_2002_c2

NAPTR

Cisco.com

- Naming Authority PoinTeR
- Universal resource identifier
- Regular expressions
- Replacement strings
- RFC 2915

Load Sharing

Cisco.com

- Resolvers use addresses in the order received, although the original concept was that they choose randomly
- DNS server can rotate the order of the (multiple) addresses of a hostname to distribute the load

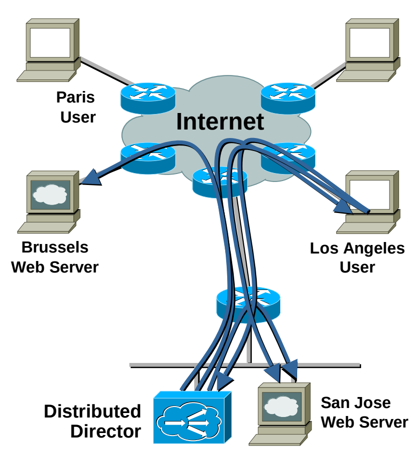
NSC-141
5043_04_2002_c2
© 2002, Cisco Systems, Inc. All rights reserved.

Source-Dependent Answers

- Return addresses in the order of "closeness" to the resolver
- Same subnet is close, but requires knowing the subnet mask
- Can look into the routing structure
- DNS support for content networking uses other metrics for which answer to give

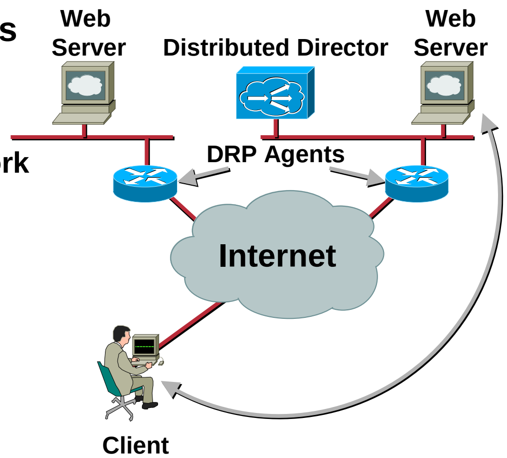
Distributed Director

- DD Server is authoritative nameserver for zone with name as webserver
- When query arrives, DD server verify with DRP agents which webserver is closest to client
- DD server respond with "best" IP-address
- DD server can also do HTTP redirects in HTTP mode



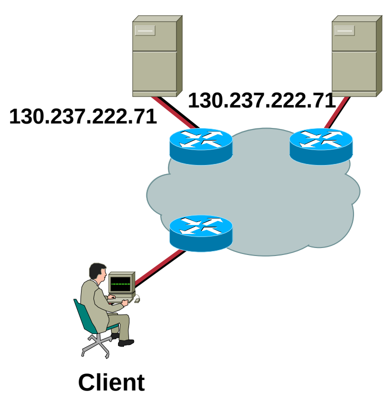
DRP—Director Response Protocol

- Operates with routers in the field to determine:
  - Client to server network proximity
  - Client to server link latency (RTT)
- UDP based

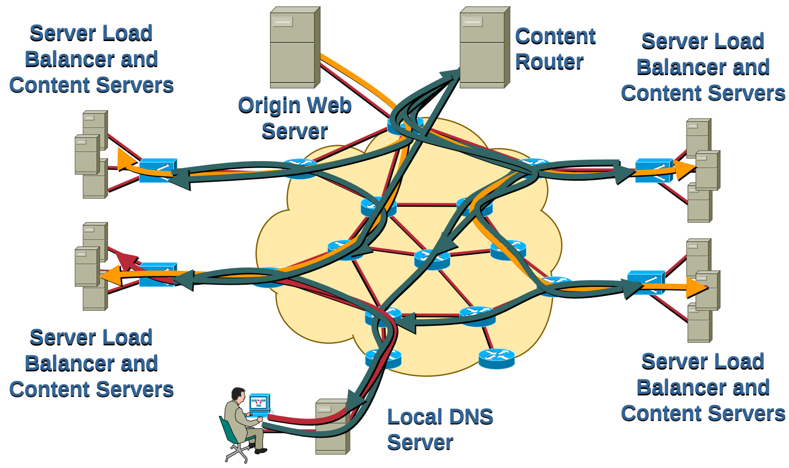


Anycast

- Announce the same IP-address from multiple servers
- Only works with UDP
- Two versions:
  - Anycast inside an AS (ok)
  - Anycast where the same AS is announced from multiple sources (dangerous due to security reasons)



Boomerang



1. Content is distributed
2. Client query content router via local DNS Server
3. Content servers are told to respond to DNS query
4. Responses are sent back
5. First response arriving is closest
6. Client connect to that IP-address

Summary

- Addresses can be allocated automatically
- DNS can support more than just name to address lookup

Cisco.com

Understanding DNS and DHCP

Session NSC-141

John Schnizlein

Cisco.com

Please Complete Your Evaluation Form

Session NSC-141

Cisco Systems

Empowering the Internet Generation

Reverse Delegation

- RFC 2317
  2.3/12 CNAME xxx.IN-ADDR.ARPA

```
For delegation of

192.0.2.0/25   to organization A
192.0.2.128/26 to organization B
192.0.2.192/26 to organization C
```


Reverse Delegation Problem

```
$ORIGIN 2.0.192.in-addr.arpa.
 ;
 1               PTR     host1.A.domain.
 2               PTR     host2.A.domain.
 3               PTR     host3.A.domain.
 ;
 129             PTR     host1.B.domain.
 130             PTR     host2.B.domain.
 131             PTR     host3.B.domain.
 ;
 193             PTR     host1.C.domain.
 194             PTR     host2.C.domain.
 195             PTR     host3.C.domain.
```


Reverse Delegation Solution

```
$ORIGIN 2.0.192.in-addr.arpa.
   @       IN      SOA     my-ns.my.domain.
hostmaster.my.domain. (...)
   ;...
   ;  <<0-127>> /25
   0/25            NS      ns.A.domain.
   0/25            NS      some.other.name.server.
   ;
   1               CNAME   1.0/25.2.0.192.in-addr.arpa.
   2               CNAME   2.0/25.2.0.192.in-addr.arpa.
   3               CNAME   3.0/25.2.0.192.in-addr.arpa.

   ;  <<128-191>> /26
   128/26          NS      ns.B.domain.
   128/26          NS      some.other.name.server.too.
   ;

   129          CNAME   129.128/26.2.0.192.in-addr.arpa.
   130          CNAME   130.128/26.2.0.192.in-addr.arpa.
   131          CNAME   131.128/26.2.0.192.in-addr.arpa.
```


Reverse Delegation Solution (Cont.)

Cisco.com

```
;  <<192-255>> /26
192/26       NS      ns.C.domain.
192/26       NS      some.other.third.name.server.
;
193          CNAME   193.192/26.2.0.192.in-addr.arpa.
194          CNAME   194.192/26.2.0.192.in-addr.arpa.
195          CNAME   195.192/26.2.0.192.in-addr.arpa.

$ORIGIN 0/25.2.0.192.in-addr.arpa.
@       IN   SOA     ns.A.domain. hostmaster.A.domain. (...)
@            NS      ns.A.domain.
@            NS      some.other.name.server.
;
1            PTR     host1.A.domain.
2            PTR     host2.A.domain.
3            PTR     host3.A.domain.
```