CS 4/55231
Internet Engineering

Kent State University
Dept. of Computer Science

LECT-7A

SCALABLE INTERNET ROUTING

Scalability

- The management of global resource is a complex task. The first set of IP level problems are:
  - IP address space depletion.
  - Optimization of route information propagation
  - Routing table explosion

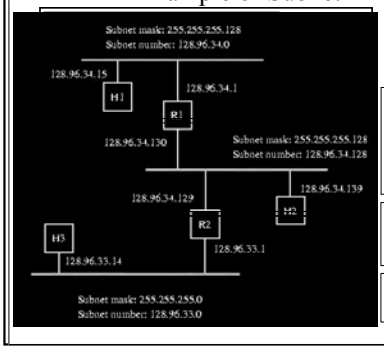
INTERNET ENGINEERING

Basic Subnetting & Subnet Masks

- IP address is depleting faster than expected:
  - All network, even one with 2 hosts, need at least class C address.
  - A network with 256 hosts need class B address.
- Also, the more there is networks, the bigger the routing table gets.
- Solution is *subnetting*.
  - A network can be divided into subnets.
  - Outside routers still view them as one large network.
  - Only, the local routers see them as separate networks.

INTERNET ENGINEERING

Example of Subnet



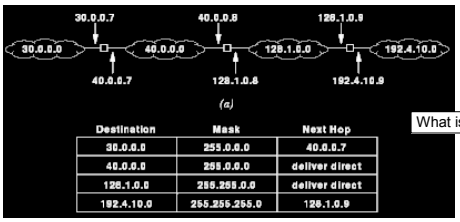
INTERNET ENGINEERING

•But the administrator of 128.96.34.0 has divided its network into two physical networks 128.96.34.0 and 128.96.34.0 and 128.96.34.128 with mask 255.255.255.128

•Network 128.96.34.0 is class B address and can have about 256x256 hosts in one large network

• H3 in 128.96.33.0 sees everything in 128.96.34.0 as one single network.

Routing Table with Subnet Masks



Destination	Mask	Next Hop
30.0.0.0	255.0.0.0	40.0.0.7
40.0.0.0	255.0.0.0	deliver direct
128.1.0.0	255.255.0.0	deliver direct
192.4.10.0	255.255.255.0	128.1.0.9

What is Bit Mask?

- R2's routing table

INTERNET ENGINEERING

IP Address Space Depletion

- The growing demand for IP addresses has put a strain on the classfull model, specially class B space, which is depleting rapidly.
- Solution Proposals
  - Creative IP address space allocation.
  - Classless Inter domain Routing (CIDR).
  - Private Addressing Network Address. Translation (NAT).
  - IPv6.

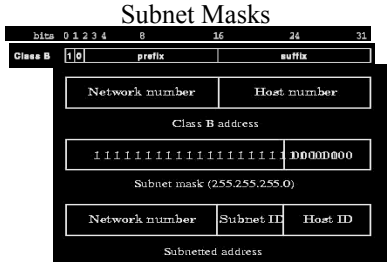
INTERNET ENGINEERING

IP Address Allocation

- Class A is very limited in number. Upper range A 64-127 will be distributed but there is still no plan on how.
- Class B addresses are also restricted. They will be allocated only if the need can be justified. Due to low utilization, the recommendation is to allocate blocks of class C addresses.
- Class C is now being distributed in way which is compatible with *address aggregation* techniques.
- Class C are being distributed to ISPs with the requirement that the original allocation for the provider should last at-least two years.
  - Each ISP must allocate half from its old address space for new customer request.
  - A customer will not be granted new address space if its current allocation is less than 80%.

INTERNET ENGINEERING

Subnet Masks



- Now in each routing table both the network number and the mask is stored.
- An AND operation is performed before looking up for the next hop.
- For distant networks, the mask is of type A, B, or C. But for local network, the mask is longer.

INTERNET ENGINEERING

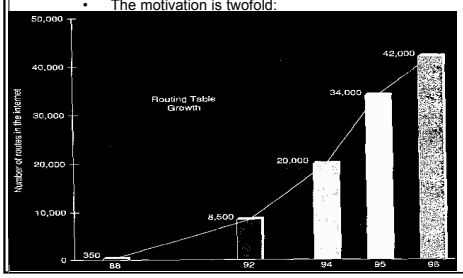
IP Address Allocation (cont.)

- A customer can receive a class-B if its requirement is more than 4096 IP addresses.
- Follow the geographical allocation of blocks in class C addresses:
  - Multiregional 192.0.0.0 to 193.255.255.255
  - Europe 194.0.0.0 to 195.255.255.255
  - Others 196.0.0.0 to 197.255.255.255
  - North America 198.0.0.0 to 199.255.255.255
  - Cen/South America 200.0.0.0 to 203.255.255.255
  - Pacific Rim 202.0.0.0 to 204.255.255.255
  - Others 204.0.0.0 to 205.255.255.255
  - Others 206.0.0.0 to 207.255.255.255

INTERNET ENGINEERING

Classless Interdomain Routing

- This is an attempt to remove the classification of IP addresses.
- The motivation is twofold:



INTERNET ENGINEERING

Classless Internet Routing

- IP address is represented as the IP and a count of leftmost contiguous bits. Example:
  - 198.32.0.0/16
  - similar to 198.32.0.0 with mask 255.255.0.0, but the length can be any number.
- "Supernet" CIDR Block and Aggregate:
  - A network is classed as **supernet** when the prefix boundary contains fewer bits than the networks natural mask.
  - 198.32.0.0/16 is a supernet to 198.32.1.0 with a natural mask 255.255.255.0.
  - CIDR block, aggregate, and supernet all refers to a specific block of contiguous IP addresses.

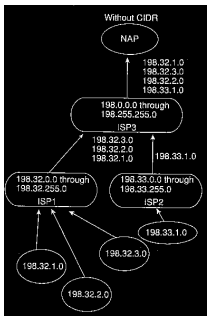
INTERNET ENGINEERING

Classless Internet Routing

- "More specific" network
  - All the networks that are a subset of an aggregate or a CIDR block are called "more specific" because they give more information about the location of a network.
  - Example: 198.213.0.0/16 -aggregate of length 16, while 198.213.1.0/20 is more specific prefix of length 20.
- CIDR Routing Domains
  - routing domains which are CIDR capable are called classless.
- Hierarchical Internet Architecture
  - CIDR provides more efficient means for building hierarchical network architecture.
  - It tremendously saves route propagation in **leaf** networks.

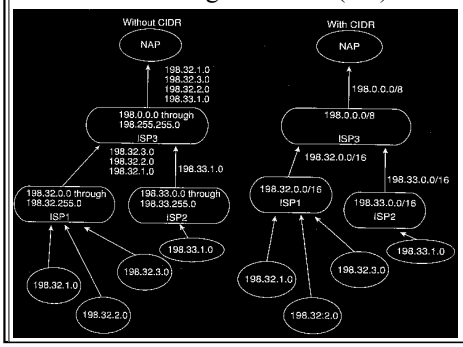
INTERNET ENGINEERING

Advantage of CIDR (1/2)



INTERNET ENGINEERING

Advantage of CIDR (2/2)



INTERNET ENGINEERING

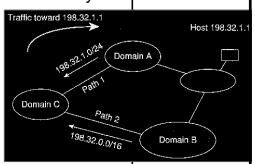
Aggregation is not Always Easy

- Aggregation works optimally if every customer is *single homed* and they takes their address from their providers address space only.
- However, this is not always the case in real world. The real world routing is much complex with CIDR.

INTERNET ENGINEERING

Routing Rules

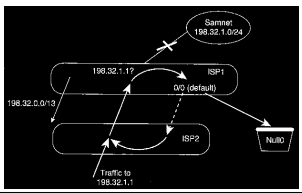
- Longest Match Rule
  - Routing to all destinations is always done on a longest match basis.
  - A router that has to decide between two different length prefixes of the same network will always follow the longer mask.
- Example:
  - A router has two matching entries:
  - 198.32.1.0/24 via path 1
  - 198.32.0.0/16 via path 2
  - Router should forward via path 1



INTERNET ENGINEERING

Routing Rule: Router Loop

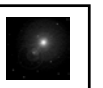
- Exception Rule
  - A network must not follow a less specific route, if that route is part of its own aggregation.
  - Use pit-bucket.



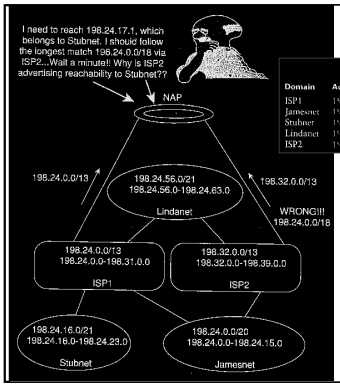
INTERNET ENGINEERING

Routing Rule: Black Hole

- A black hole occurs when traffic reaches and stops at destination that is not its intended destination, and it cannot be forwarded out of it.
- It Can occur:
  - Single Homing Scenario: when a customer network borrows IP address from address space which is not part of its providers address space.
  - Multi Homing Scenario: when a customer is connected via two providers but has IP address from one.



INTERNET ENGINEERING

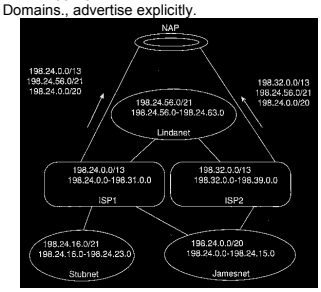


Domain	Address Range	Aggregate	Provider	Add Take Fron
ISP1	198.24.0.0-198.31.0.0	198.24.0.0/13		
Jamesnet	198.24.0.0-198.24.15.0	198.24.0.0/20	ISP1, ISP2	ISP1
Stubnet	198.24.16.0-198.24.23.0	198.24.16.0/21	ISP1	ISP1
Lindanet	198.24.56.0-198.24.63.0	198.24.56.0/21	ISP1, ISP2	ISP1
ISP2	198.32.0.0-198.39.0.0	198.32.0.0/13		

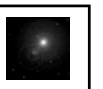
- ISP2's is aggregating Lindanet and Jamesnet's addresses and advertising the aggregate. This aggregation, unfortunately erroneously implies that Stabnet as well can be reached from ISP2.

Routing Rule: Black Hole

- Never aggregate the address spaces of other Routing Domains., advertise explicitly.

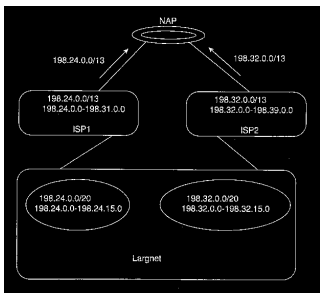


- Costing? ISP1 or ISP2, who is going to receive the traffic for Jamesnet and Lindanet?



INTERNET ENGINEERING

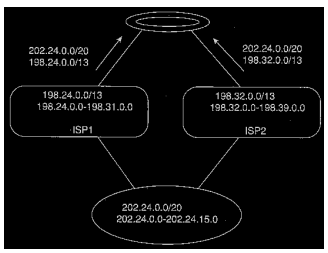
Problem: Partial Reachability



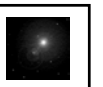
- In many situation a large network will have IP addresses from more than one ISPs. In that case both ISPs must explicitly advertise the address space allocated by the other to its customer. Otherwise, there will be partial outage.

INTERNET ENGINEERING

Problem: Irregular IP allocation



- The customer has IP address which belong to none of the ISPs attached to it. This will prevent aggregation.



INTERNET ENGINEERING

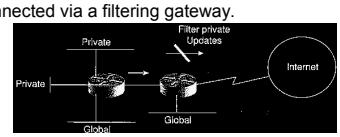
Aggregation Recommendations

- A domain that has been allocated a range of addresses has the sole authority for aggregation of its address space.
- When a domain performs aggregation, it should aggregate as much as possible.
- Different situation requires different design. No one solution can handle all situation.
- The introduction if CIDR has helped dampening the explosion of global routing tables.
- BGP4 is the routing protocol of choice that helps efficient handling of route aggregation and propagation.

INTERNET ENGINEERING

Private Addressing

- Global vs. Local connectivity.
  - ATM, cash registers etc. may never require global connectivity.
- IANA allocated few addresses for local use:
  - 10.0.0.0-10.255.255.255 (one class A network)
  - 172.16.0.0-172.31.255.255 (16 class B networks)
  - 192.168.0.0-192.168.255.255 (256 class C networks)
- These addresses can be reused provided they do not communicate to the outside world, and they are connected via a filtering gateway.

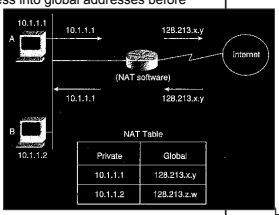


INTERNET ENGINEERING

Network Address Translation

- NAT routers
  - placed at the border of a domain
  - enables private networks to connect to the Internet without resorting to tedious renumbering of the hosts.
  - translates the private address into global addresses before sending packets out.

- A and B have IP addresses 10.1.1.1 and 10.1.1.2. If A and B want to reach outside the router will automatically change their IP addresses into a valid global IP 128.21.3.x.y. A smaller pull of IPs can be reused.



INTERNET ENGINEERING

IP v6

- IP V6 allocates 128 bits (compared to 32 bits in IPv4) to address the problem of IP address depletion.
- IPV6 also has also definition for multiple classes of addresses. It is indicated by a variable length FP (Format Prefix) field.

Description	Format Prefix
Reserved	0000 0000
Unassigned	0000 0001
Reserved for NSAP Allocation	0000 001
Reserved for IPX Allocation	0000 010
Unassigned	0000 011
Unassigned	0000 1
Unassigned	0001
Unassigned	001
Provider-Based Unicast Address	010
Unassigned	011
Reserved for Geographic Unicast Addresses	100
Unassigned	101
Unassigned	110
Unassigned	1110
Unassigned	1111 0
Unassigned	1111 10
Unassigned	1111 110
Unassigned	1111 1110 0
Link Local Use Addresses	1111 1110 10
Site Local Use Addresses	1111 1110 11
Multicast Addresses	1111 1111

INTERNET ENGINEERING

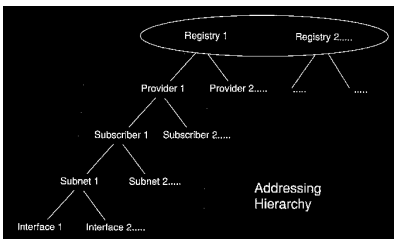
IP v6

- A provider based unicast address has FP =010. It is subdivided into following fields:
  - REGISTRY ID
  - PROVIDER ID
  - SUBSCRIBER ID
  - SUBNET ID
  - INTERFACE ID (it can be MAC address as well).

3	x bits	y bits	z bits	w bits	125-x-y-z-w bits
010	REGISTRY ID	PROVIDER ID	SUBSCRIBER ID	SUBNET ID	INTERFACE ID

INTERNET ENGINEERING

IP v6



3	x bits	y bits	z bits	w bits	125-x-y-z-w bits
010	REGISTRY ID	PROVIDER ID	SUBSCRIBER ID	SUBNET ID	INTERFACE ID

INTERNET ENGINEERING

Link Local and Site Local IP

- Link local addresses are private to a particular physical network.
- Site local addresses are private to a particular site.

10 bits	x bits	118-x bits
1111111010	00000..	INTERFACE ID

Link Local-Use Addresses

10 bits	x bits	y bits	118-x-y bits
1111111011	00000..	SUBNET ID	INTERFACE ID

Site Local-Use Addresses

- Quiz: What if later on the company decides to make these addresses global?

INTERNET ENGINEERING

Internet Interconnection Structure

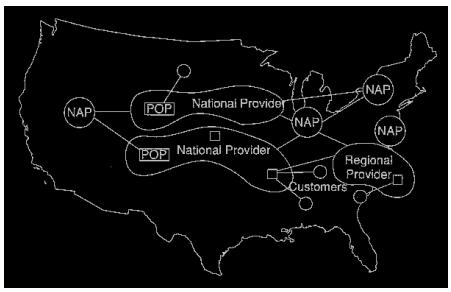
31

Basic Concepts (1)

INTERNET ENGINEERING

- Internet Service Provider (ISP)
  - Provider who connects an end user customer with the Internet in one or few geographic regions.
- National & Regional Provider (NP and RP)
  - Provider who connects two or more ISP networks across regions.
- Point-of-Presence (POP)
  - An access point where a customer can connect into an ISP network.

Concepts (POP, ISP, NP)



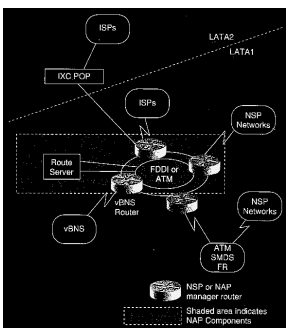
INTERNET ENGINEERING

Basic Concepts (2)

INTERNET ENGINEERING

- National Access Point (NAP)
  - The facility where various NPs networks can interconnect.
  - Formerly these were organized as Federal Internet Exchange (FIX) & Commercial Internet Exchange (CIX). FIX/CIX model did not scale well.
  - It is physically a high-speed network switch or network to which a number of routers can be connected for the purpose of traffic exchange (example: FDDI or ATM switch).

NAP



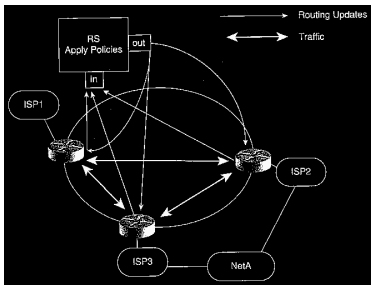
INTERNET ENGINEERING

Basic Concepts (3)

INTERNET ENGINEERING

- Route Server
  - Route server exchanges routing information and policy with the service provider routers attached to the NAP.
  - It does not perform any traffic forwarding.
  - A group of servers facilitates interconnections between ISPs by gathering routing information from each ISP applying ISPs predefined set of rules, policies, and then redistributing the processed information to each ISP.
  - It saves routers of each individual ISPs to peer with all other routers, thus cutting down the number of peers from (n01) to 1.

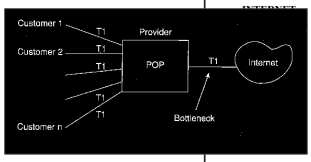
Route Server

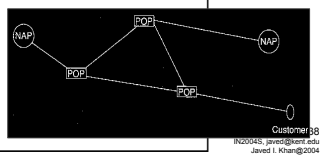


INTERNET ENGINEERING

ISP-Customer Connectivity (1)

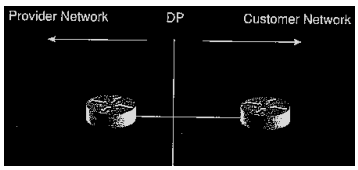
- ISP Bottleneck
  - Each ISP should have ample bandwidth to provide service to all its customers.
- Access Redundancy
  - ISP connected via multiple NAP will be more reliable than the one with one.





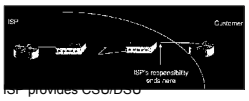
ISP-Customer Connectivity (1)

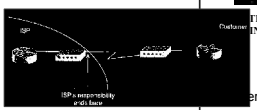
- Demarcation Point
  - A demarcation point is the point that differentiates a curtomers network from that of the provider.



INTERNET ENGINEERING

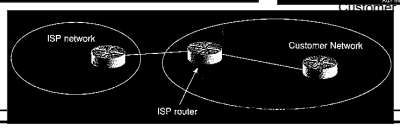
Demarcation Models





customer provides router

Customer router at providers site



INTERNET ENGINEERING

Autonomous Systems

Autonomous System

- Autonomous Systems
  - An autonomous system (AS) is a set of routers having a single routing policy, running under a single technical administration. Each AS has its own internal routing mechanism and policy. Each AS has a unique number registered at IANA.

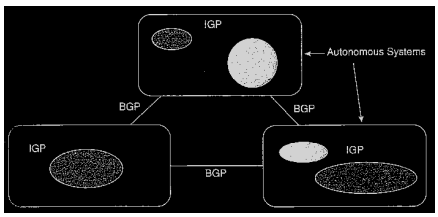
INTERNET ENGINEERING

Divide the Routing

- Interior Routing Protocol
  - The routing protocol(s) that an AS use internally to route its packets are knows as Interior Gateway Protocols or IGP.
  - Examples: RIP, OSPF
- Exterior Routing Protocol
  - The routing protocol that the AS use to exchange traffics between them is called Exterior Gateway Protocol or EGP.
  - Example: BGP

INTERNET ENGINEERING

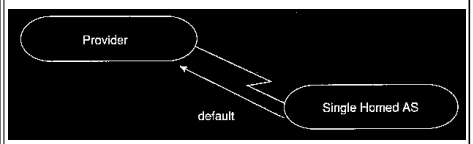
IGP and BGP



INTERNET ENGINEERING

AS Connectivity

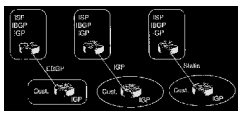
- Stab Autonomous System
  - An Autonomous System is called a stub, when it reaches networks outside its domain via a single exit point. These are also called single homed AS.
  - A Stab AS does not really have to learn Internet routes from its provider. All traffic can be default routed to provider.



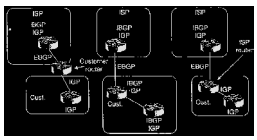
INTERNET ENGINEERING

Routing Alternatives for Stab AS

- Static, IGP, or EGP:



- Options with a Customer Router



INTERNET ENGINEERING

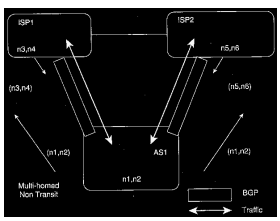
Multi-homed AS

- Multi-homed Non-transit AS
  - An AS is multi-homed if it has more than one exit point to the outside world. A non-transit router does not allow traffic of of any source or destination that is outside the AS.
  - Like single homed routers, here also EGP is not required between AS1 and providers.. However, is recommended.

INTERNET ENGINEERING

Multi-homed AS

- Multi-homed Non-transit AS
  - It will advertise only its own routes, not the one learned from the other ASs.
  - Example: AS1 only advertises n1 and n2.



- ISP1 and ISP2 still can force packets in AS1 by static route.
- AS1 can however, filter packets.

INTERNET ENGINEERING

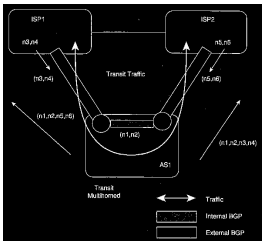
Multi-homed Transit AS

- Multi-homed Transit AS
  - A multi-homed transit router is connected to more than one provider. It allows traffic with origin and destination that does not belong to it.
- Some additional concepts:
  - Border Router: Routers those run EGP
  - IBGP: For transit traffic an AS runs an internal version of EGP, it is called **Interior Border Gateway Protocol** (IBGP).
  - Transit Router: Routers those run IBGP.
  - For handling transit traffic an AS creates 'pipe' between two border routers who are also transit routers.

INTERNET ENGINEERING

Multi-homed Transit AS

- For handling transit traffic an AS creates 'pipe' between two border routers who are also transit routers.



INTERNET ENGINEERING

Next Topic:
BGP4