Thoughts on Address Prefix Management

Fred Baker

Cisco Fellow

 © 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Problem:

- Marla Azinger (ARIN) would like for the IETF to give advice to ISPs on how to filter IPv6 prefixes in BGP

  IETF's idea: filter to allocation levels (/32)

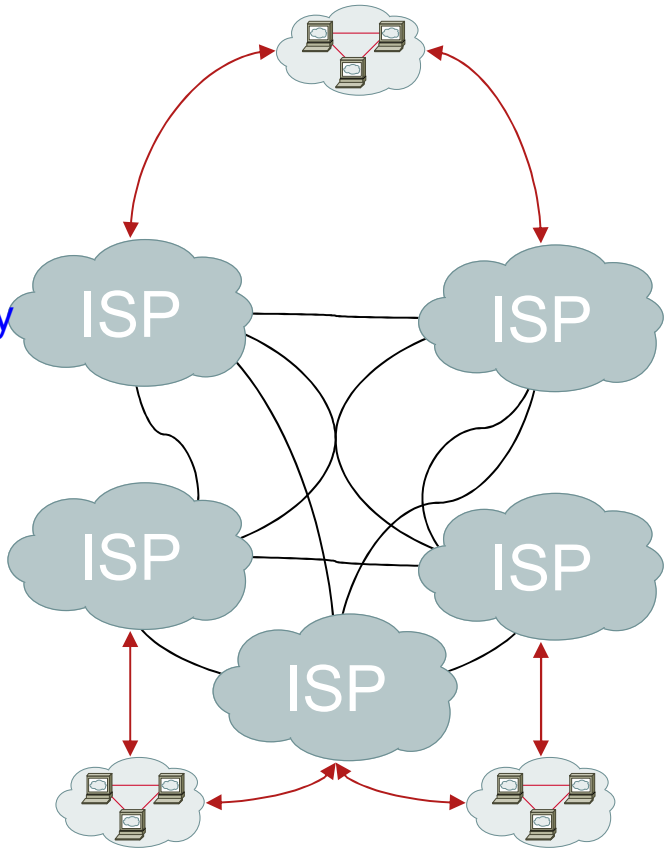
  Marla's idea: filter to assignment levels, which is to say individual edge networks (/48, /56).

- IETF IPv6 Operations WG discussion, in design team

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Multihoming

- Principal reason for edge network getting an AS number is multihoming:
  - Network viewpoint:
    - Prefixes are assigned to entities whose routing connectivity and size make it advantageous to maintain global knowledge of their routing and who desire additional guarantees of internet connectivity
  - Customer viewpoint:
    - Service is obtained from multiple providers to improve reliability or other characteristics
- Two general forms:
  - Provider-independent (PI)
    - Prefix assigned to edge network forever
  - Provider-assigned/Provider-aggregatable (PA)
    - Prefix assigned to edge network by ISP and advertised through multiple ISPs



© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

RFC 3582 multihoming requirements

Redundancy	Shields edge from network failures
Address portability	ISP-portable Prefixes
Load sharing	Controlled by edge
Performance	Traffic distributed by edge policy
Policy	Edge network can use any policy
Simplicity	Simple to install/maintain
Transport session survivability	Sessions survive failures
Impact on DNS	No DNS impact
Datagram filtering	Not affected by ISP ingress filtering
Scaling: impact on routers	Route table prefix count
Scaling: impact on hosts	Requires no host changes
Scaling: host/router interaction	No change to Neighbor Discovery etc
Scaling: network management	Simple to monitor/configure
Scaling: ISP cooperation	Requires no ISP cooperation

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

What does it mean for addressing to "scale"?

- Protocols and procedures are said to "scale" when they
  - Operate well on all deployment scales, including global
  - Manage growth with no proportional increase in cost or effort, and preferably proportionally decreasing effort
- Conclusion:
  - In 2050, the planet's population will be 10,000,000,000
  - The most "scalable" address distribution architecture will minimize the number of prefixes advertised globally as compared to other approaches

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Present model - general routing

- Presently about 23,000 autonomous systems in the IPv4 network
  - There are a few hundred very large ISPs
  - 9680 advertise only one prefix; e.g., are very likely multihomed edge networks
  - The rest are split between larger multihomed edge networks, and smaller ISPs
    - http://www.cidr-report.org/, http://bgp.potaroo.net/
- Total of about 200,000 prefixes in IPv4 backbone representing those networks

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Present model - PI/PA multihoming

- Current statistics:

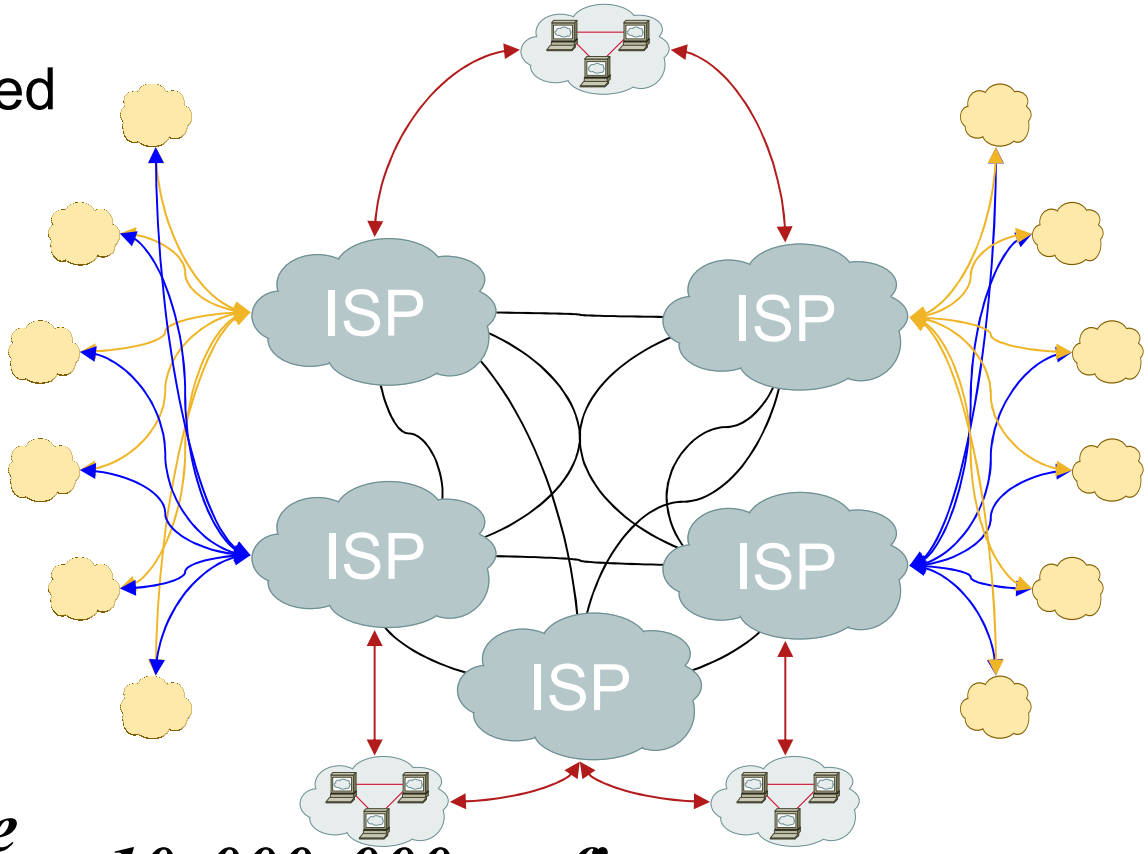
  US: about one multihomed network per 18,000 population

  World: about 1:50,000

- Expected 2050 density:

  About 1:1000?

- Implication:



$$\frac{10,000,000,000\ people}{1000\ prefixes/capita} \approx 10,000,000\ prefixes$$

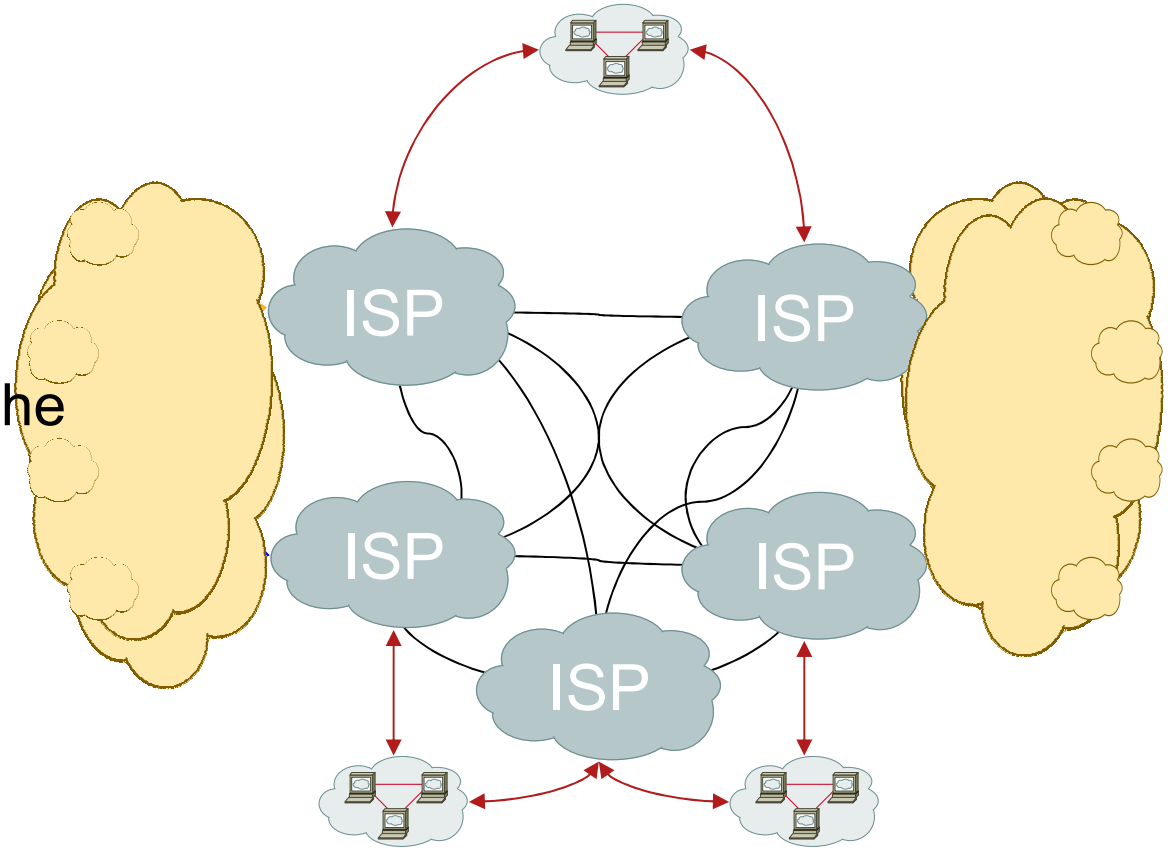
© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

RFC 3582 analysis of PI/PA multihoming

	PI	PA like PI
Redundancy	✓	✓
Address portability	✓	no
Load sharing	✓	✓
Performance	✓	✓
Policy	✓	✓
Simplicity	✓	✓
Transport session survivability	✓	✓
Impact on DNS	✓	✓
Datagram filtering	✓	Issues
Scaling: impact on routers	$O(10^7)$ prefixes	$O(10^7)$ prefixes
Scaling: impact on hosts	✓	✓
Scaling: host/router interaction	✓	✓
Scaling: network management	✓	✓
Scaling: ISP cooperation	✓	Issues

Shim6 viewpoint: PA multihoming

- Premise:
  - ISPs have prefixes
  - Edge networks inherit prefixes from ISPs
  - Only the ISP's prefix is advertised in BGP, not the inherited network prefix
- Prefixes in the internet core:
  - O(tens of thousands of prefixes)



© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

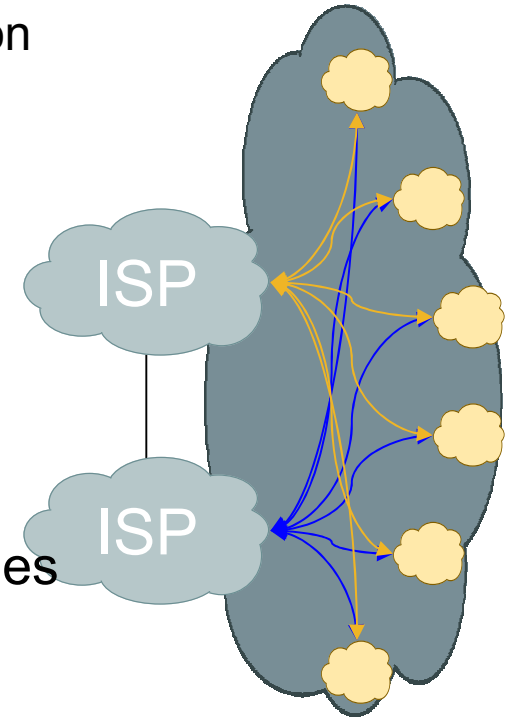
RFC 3582 analysis of shim6 multihoming

Redundancy	Multiple routes
Address portability	Addresses not portable
Load sharing	Host picks route by address pair
Performance	Performance only partially predictable
Policy	Address Pair policy is local
Simplicity	Not as simple as a single prefix
Transport session survivability	SCTP survives; UDP/TCP does not
Impact on DNS	✓
Datagram filtering	Ingress filtering affects routes
Scaling: impact on routers	O(10e^4) prefixes
Scaling: impact on hosts	Hosts must select address pair
Scaling: host/router interaction	✓
Scaling: network management	Choice of address pair not controlled in network routing but in host
Scaling: ISP cooperation	✓

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Proposal: exchange-based multihoming

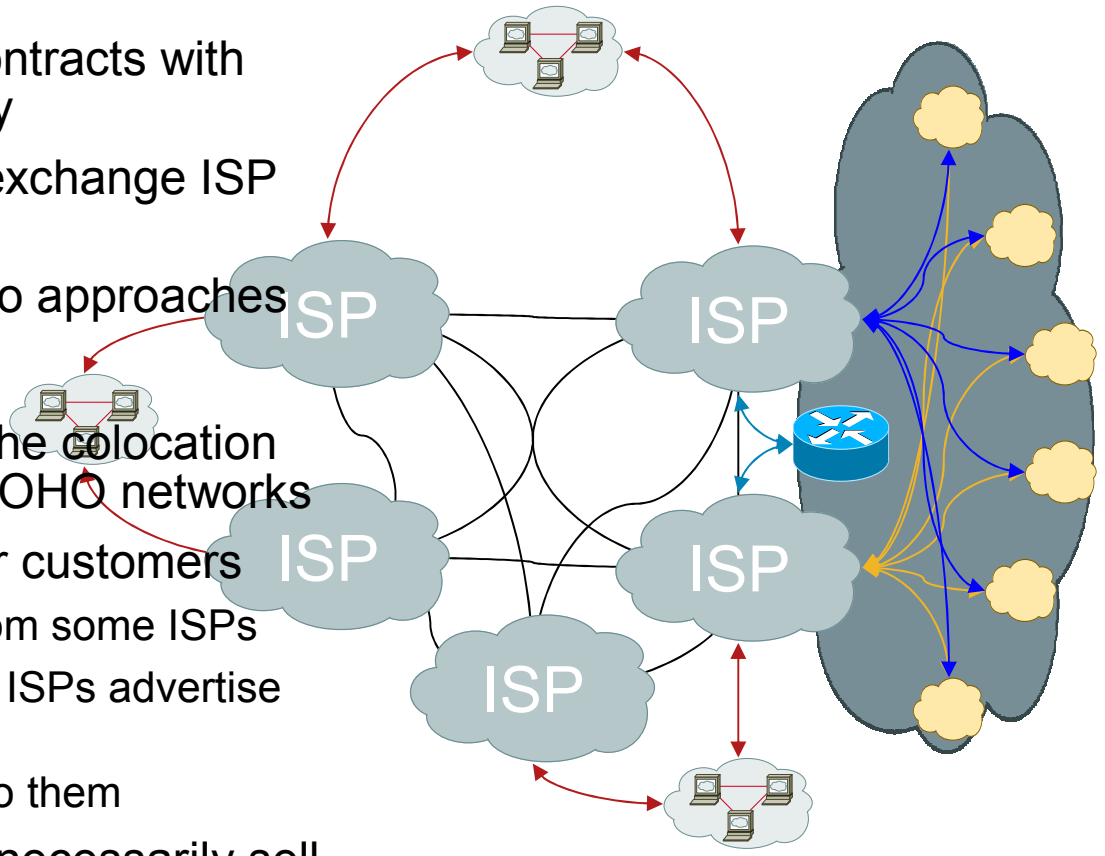
- Imagine:
  - A region that is large enough to be served by a colocation center and several ISPs, and small enough to be useful in internet routing;
    - A city or part of a large city might be an example
  - We define some regional authority such as an interchange exchange
- The exchange:
  - Allocates a prefix to the region
  - Assigns small (/64, /60, /56) prefixes to smaller entities in the region
  - Obtains agreements from the ISPs to use those prefixes for their multihomed customers and route among themselves for other customers
  - Only the larger prefix is advertised outside the region



© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Possible implementations

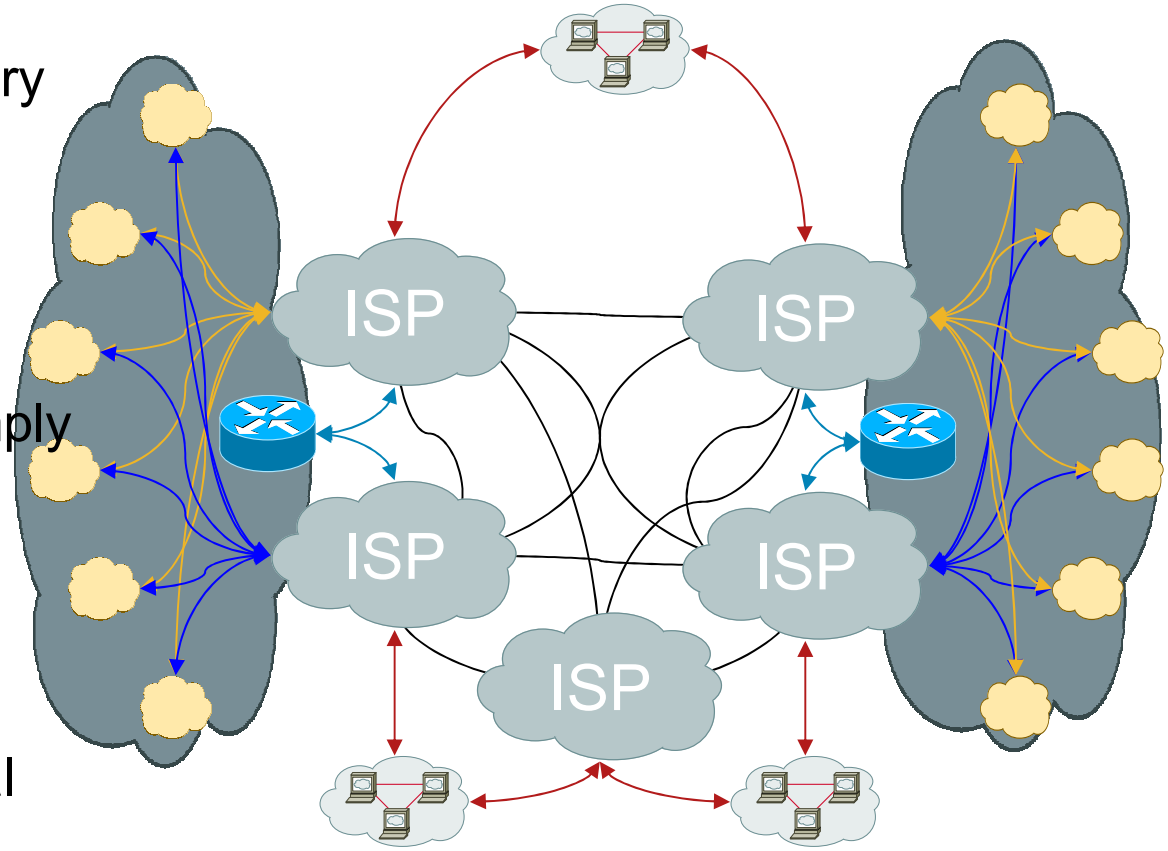
- Three obvious approaches:
  - All the ISPs maintain bilateral contracts with each other and route accordingly
  - All of the ISPs contract with an exchange ISP operated by the exchange
  - Some combination of the first two approaches
- Exchange mini-ISP model:
  - Exchange manages a router in the colocation center and assigns prefixes to SOHO networks
  - All ISPs connect to it and to their customers
    - ISP peers with or buys transit from some ISPs
    - Other ISPs buy transit from it All ISPs advertise their regional routes to it
    - It advertises the regional prefix to them
  - Note that the mini-ISP does not necessarily sell transit service outside the region
  - ISPs route directly to their customers and otherwise to the exchange ISP



© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Proposed model - exchange-based multihoming

- Imagine:
  - We deploy a prefix for every 1,000,000 people in a regional prefix
    - (Exact number not algorithmically important)
  - Interchange ISP could be government-related or simply an exchange cooperative
- The prefix identifies the general region
  - Delivery is to an ISP's customer or to the regional switch and then to the customer
- Implication:



$$\frac{10,000,000,000 \text{ people}}{1,000,000 \text{ people/exchange}} \approx 10,000 \text{ prefixes}$$

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Total prefix count in the backbone:

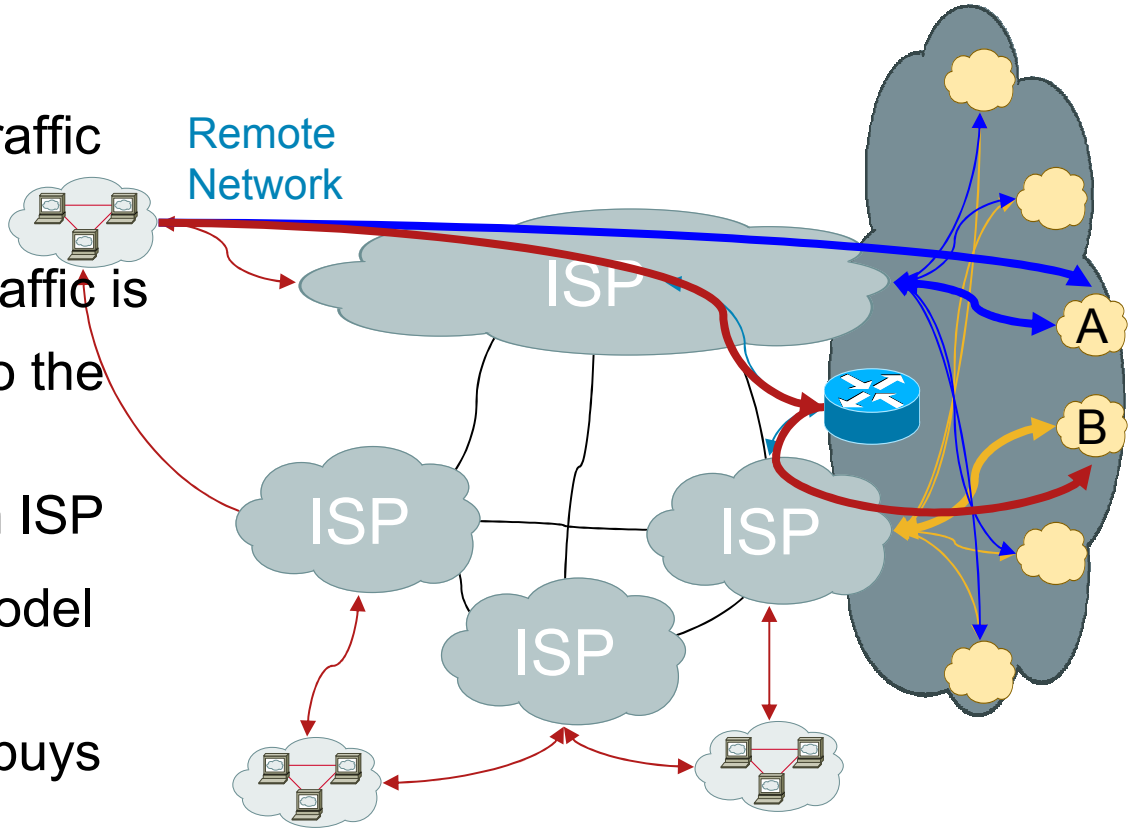
- PI/PA model:
  - One+ per ISP
  - One+ per large edge network that functions like an ISP
  - One+ per small edge network that is willing to spend the money
- Prefixes in the internet core:
  - O(tens of millions)

- Exchange-based model
  - One+ per ISP
  - One+ per large edge network that functions like an ISP
  - One+ per exchange for the smaller homes and businesses it serves
- Prefixes in the internet core:
  - O(tens to hundreds of thousands)

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Business implications of exchange-based multihoming

- Traffic is now carried by the destination's ISP
  - Hot potato routing shifts traffic there
- In exchange-based model, traffic is
  - Carried by sender's ISP to the region, and then
  - Transits to the destination ISP
- There is an implied transit model that has to be accounted for
  - Anti-trust issue: new ISP buys transit from all others?
  - Transit contracts required between exchange and carriers?



© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

RFC 3582 analysis of exchange-based multihoming

Redundancy	✓
Address portability	Portable within domain
Load sharing	✓
Performance	✓
Policy	✓
Simplicity	✓
Transport session survivability	✓
Impact on DNS	✓
Datagram filtering	✓
Scaling: impact on routers	$O(10^4 - 10^5)$ prefixes
Scaling: impact on hosts	✓
Scaling: host/router interaction	✓
Scaling: network management	✓
Scaling: ISP cooperation	Some form of exchange required

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Recommendations

- In general, ISPs should advertise and filter prefixes to allocation boundaries (/32, /48 common)
- ISPs and registries should enable peers to filter prefixes accurately by advertising rules (“prefixes are generally /32; this /32 is further sub-allocated as /48 PI”)
- In specific cases, business considerations will override, such as advertising a more specific prefix under contract.
- In such cases, they should enable peers to filter prefixes and traffic accurately
- The ISP and registry community should consider exchange-based addressing as a strategy

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Q&A

© 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Thoughts on address prefix management

Fred Baker

Cisco Fellow

 © 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential