Multihoming: An Overview

& a brief introduction to GSE(8+8)

Lixia Zhang

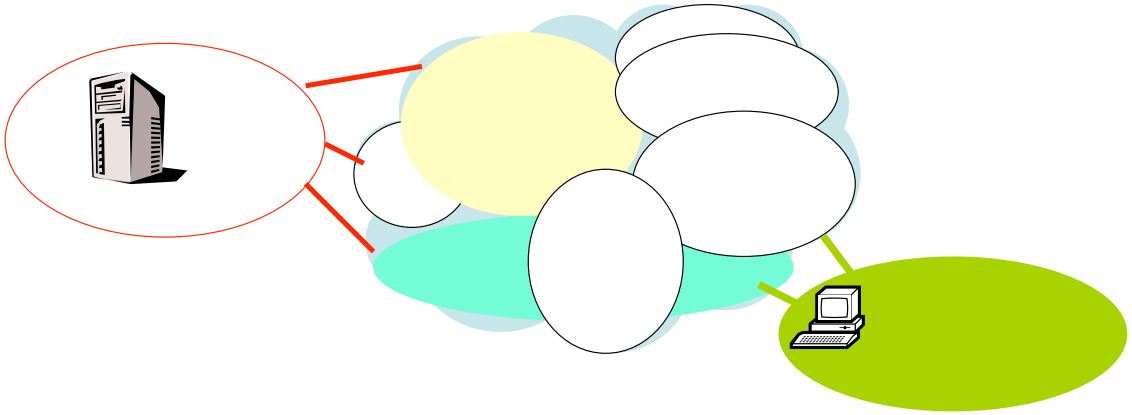
APRICOT 2006

Perth, Australia

Single Home

- Get a slice of address from your ISP
- There is only one way to send, and one way to receive data
- ISPs inject shorter (aggregated) prefixes into global routing
- Life is simple

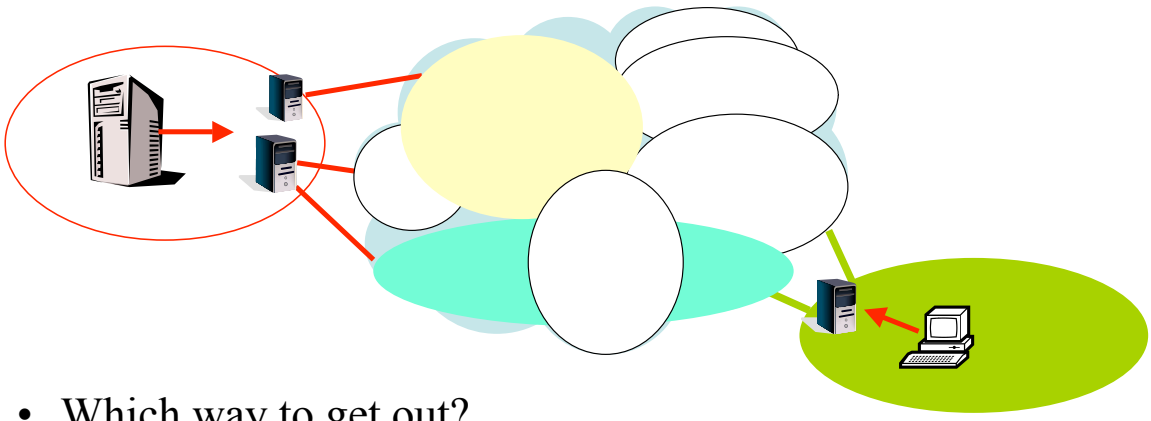
The Internet is made of multiple ISPs !



- For better reliability, cheaper price, higher throughput....
- Connect to multiple ISPs!
  - Get the best from each and all
- Bringing more questions to life ...

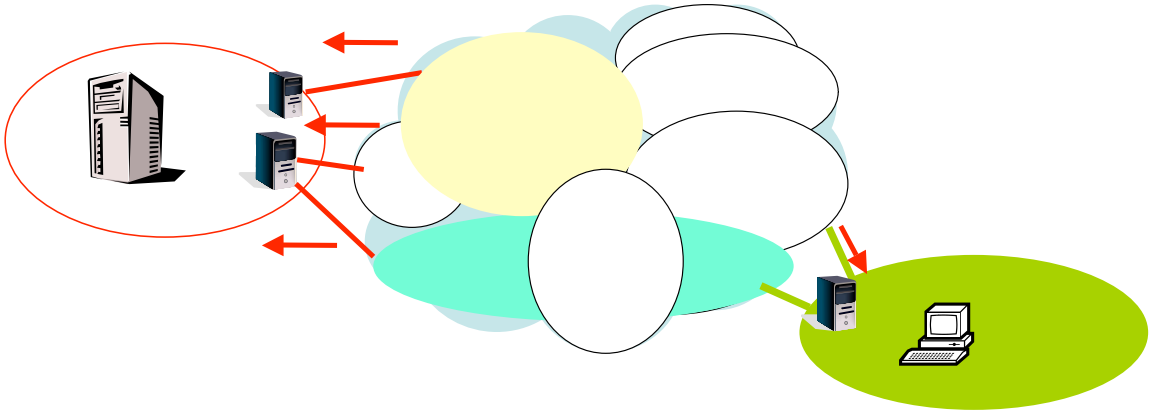
Multihoming

Multihoming: Sending Data



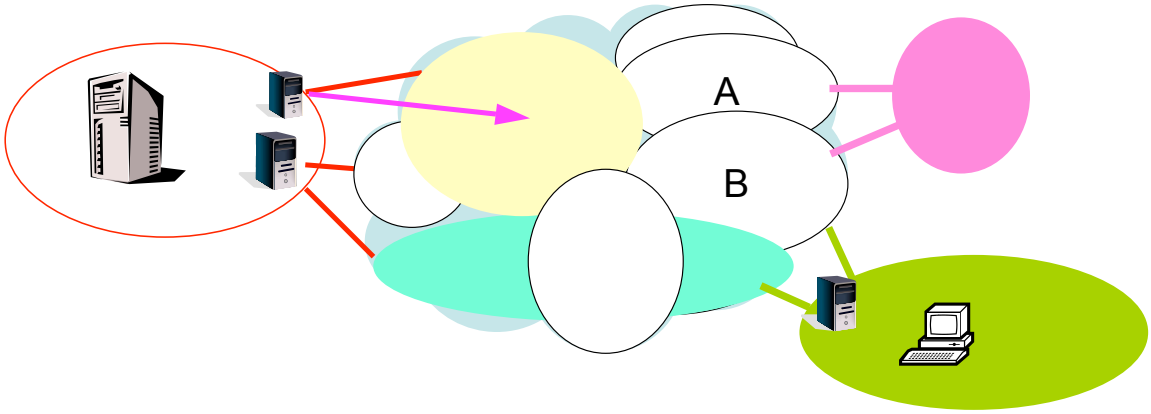
- Which way to get out?
  - You probably pay different price for different ISPs
  - The best choice may depend on specific destinations you are sending to
  - If more than one exit router: decision made *inside the site*
- *Outbound traffic engineering*

Multihoming: Receiving Data



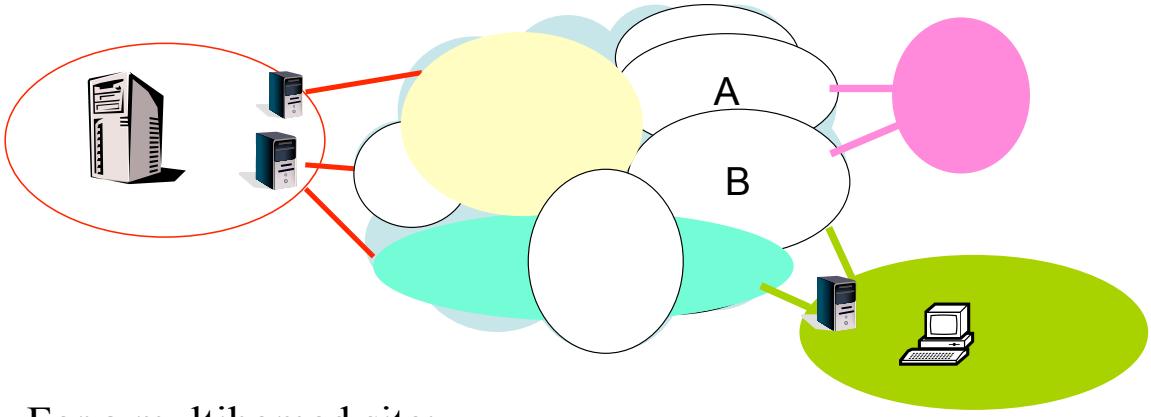
- You'd like
  - Traffic coming from a cheap link as long as it works, or
  - you'd like to utilize all links in certain proportion
- *Inbound traffic engineering*

Multihoming: lets not forget the middle



- When a ***destination*** is more than one AS hop away, a transit ISP may wish to know whether there are multiple ways to reach ●, so that it can choose the *"best"* one
- *Transit traffic engineering ™*

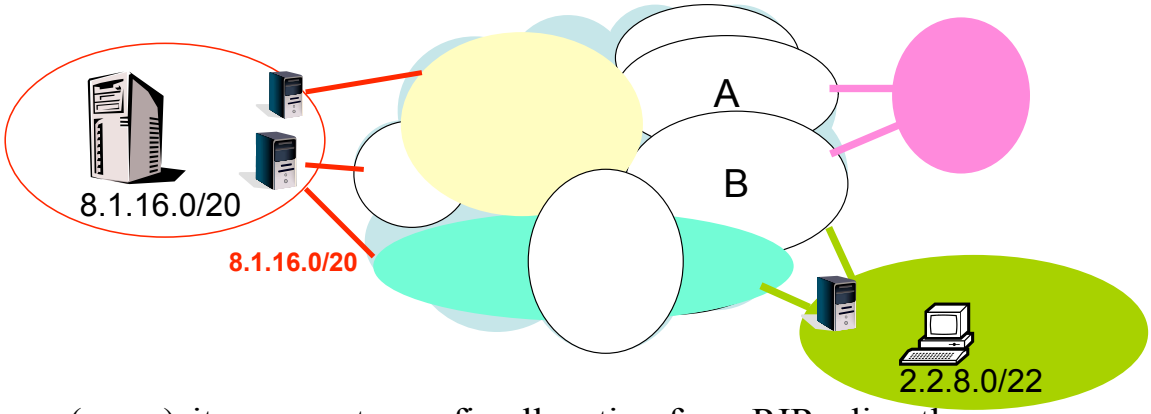
Multihoming: Addressing



For a multihomed site:

- Where does the address come from?
- What/which address to use for
  - source address?
  - destination address?

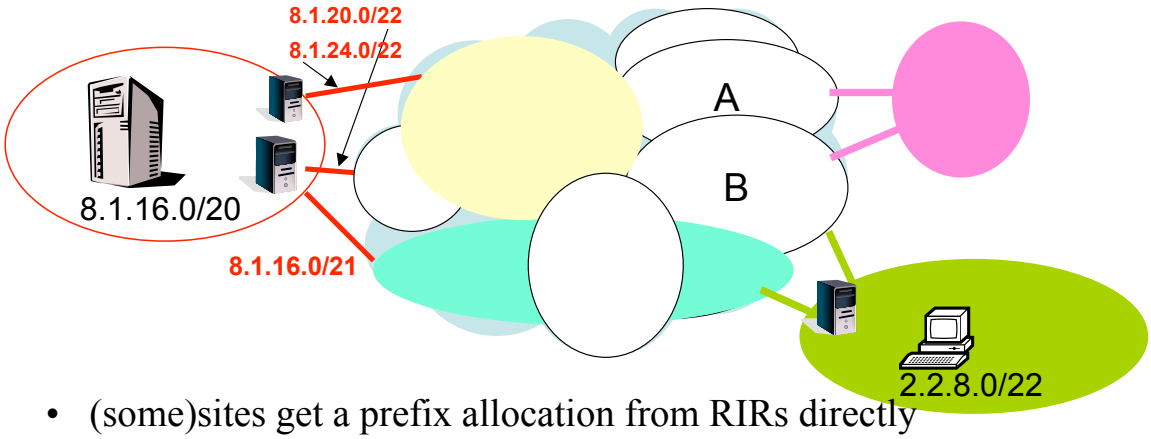
Multihoming Address: PI Prefix



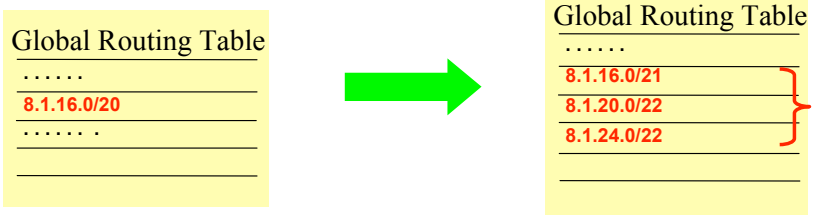
- (some)sites can get a prefix allocation from RIRs directly
  - *Every* PI prefix adds an entry into the global routing table

Global Routing Table
. . . . . .
8.1.16.0/20
. . . . . .

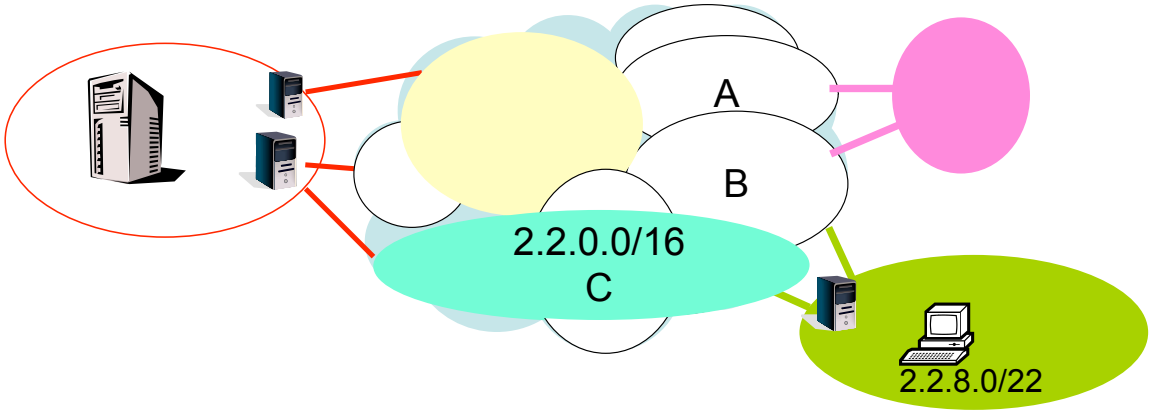
Multihoming Address: PI with TE



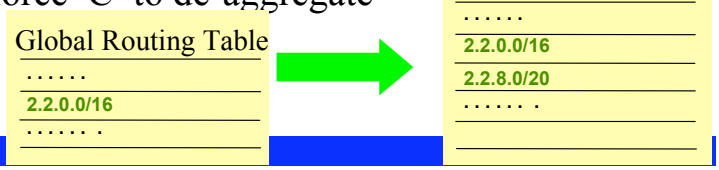
- (some)sites get a prefix allocation from RIRs directly
  - Inbound TE ⇨ split prefixes ⇨ *multiple* entries into global routing



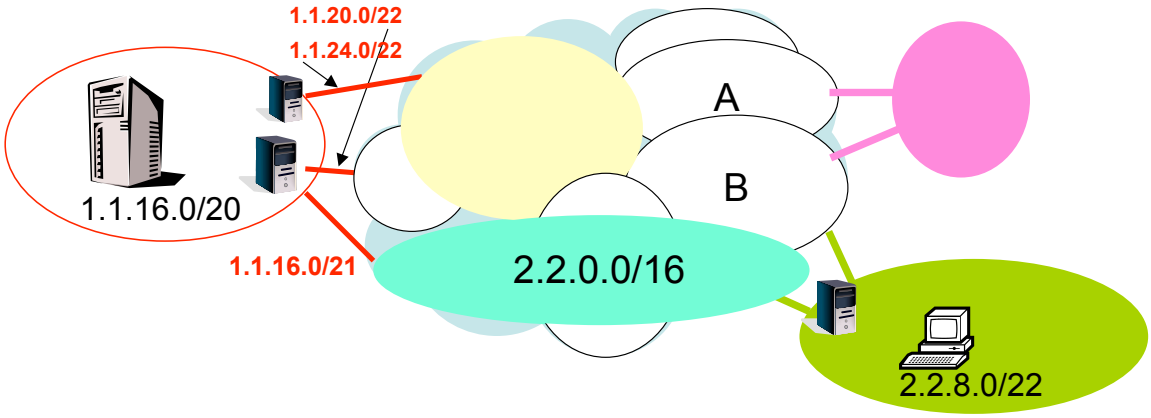
Multihoming Address: PA



- Using the prefix received from provider C
- Provider B may have to announce the prefix
  - Which in turn may force C to de-aggregate



Multihoming Address: PA + hole + split



Global Routing Table
.....
1.1.0.0/16
2.2.0.0/16
..... .

Global Routing Table
.....
1.1.0.0/16
1.1.16.0/21
1.1.20.0/22
1.1.24.0/22
2.2.0.0/16
2.2.8.0/22

Multihoming: Today's Practice

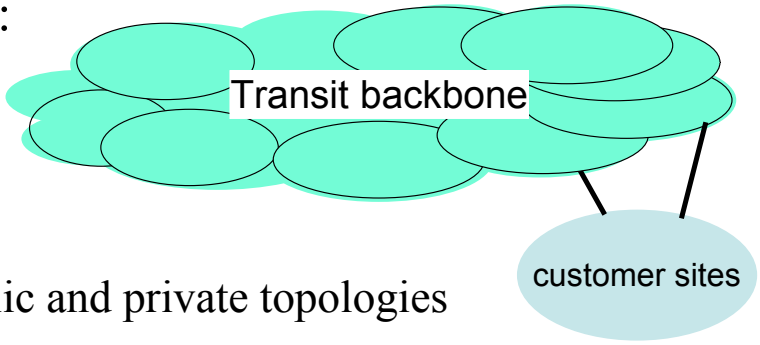
- Use of PI prefixes, or punching holes on PA prefixes
- Works as far as TE is concerned (only)

> these days, many people take TE to mean "break up my portable address block in small parts". That is one way to do it, and an effective one, but also the least scalable one --- Iljitsch van Beijnum 2/24

- Lots worries about impact on routing scalability
  - Goes up with the number of multihomed sites
  - Goes up further with additional prefix split for inbound TE

One Earlier Proposal: GSE

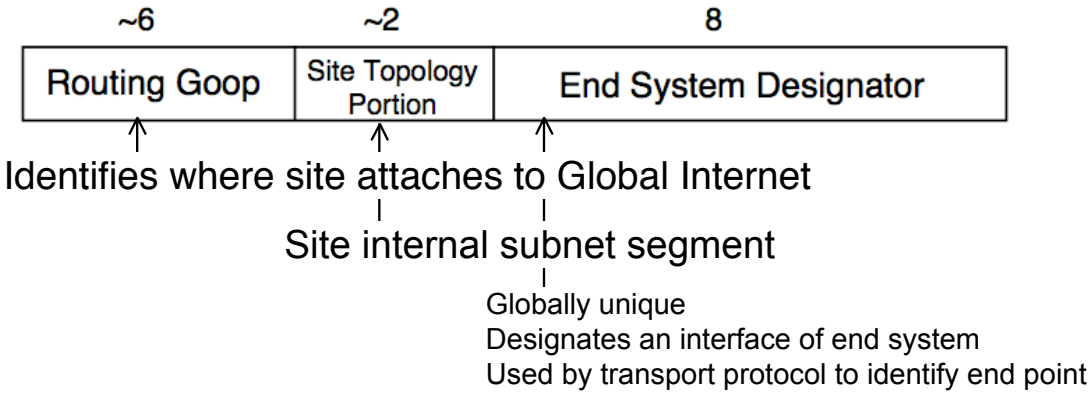
- GSE: Global, Site, and End-system address elements, proposed by Mike O'Dell in 1996-97
  http://www.watersprings.org/pub/id/draft-ietf-ipngwg-gseaddr-00.txt
- The basic idea:



  - *Separate* public and private topologies
  - *Insulate* customer site from the global provider topology

How GSE Works

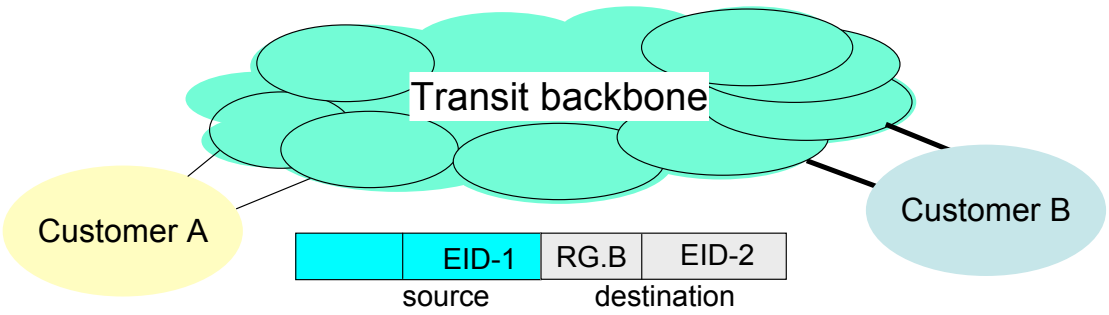
Proposed IPv6 address structure:



- Internal packet delivery without RGs
- External packet delivery: defer/hide RG AMAP
- Multihomed sites get multiple RGs

How GSE Works: more detail

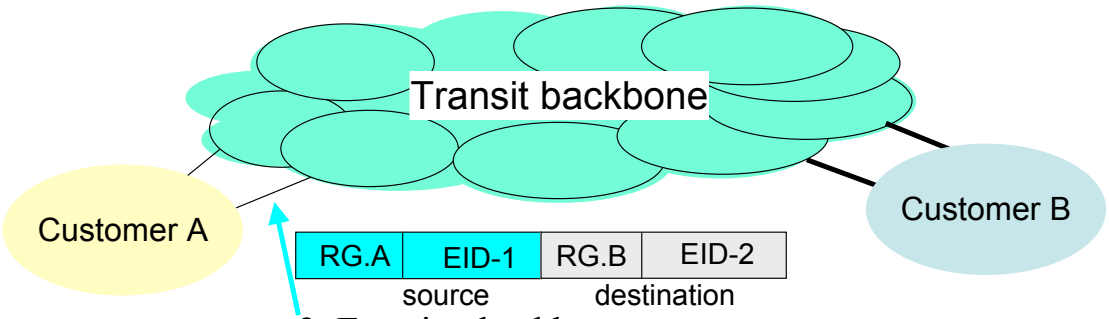
- For outbound traffic:
  - Get destination address and RG from DNS lookup
  - Put on source RG when packets *exiting* local site



1. Inside A

How GSE Works: more detail

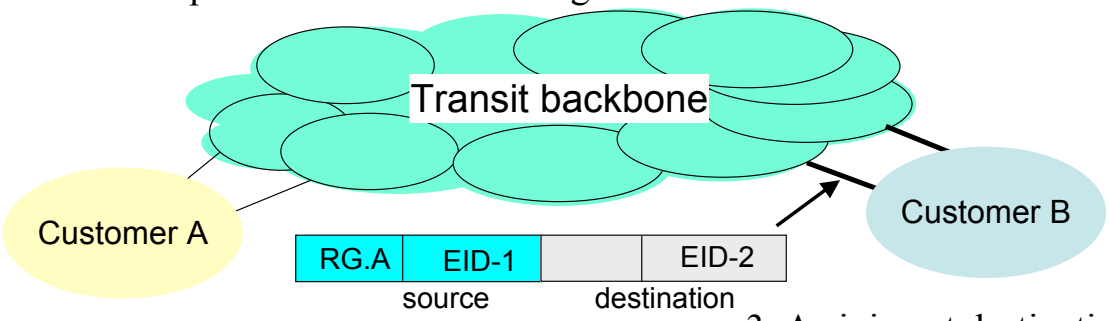
- For outbound traffic:
  - Get destination address and RG from DNS lookup
  - Put on source RG when packets *exiting* local site



2. Entering backbone

How GSE Works: more detail

- For outbound traffic:
  - Get destination address and RG from DNS lookup
  - Put on source RG when packets *exiting* local site
- For inbound traffic
  - Take off destination RG at entrance to destination site
  - Keep source RG for returning traffic



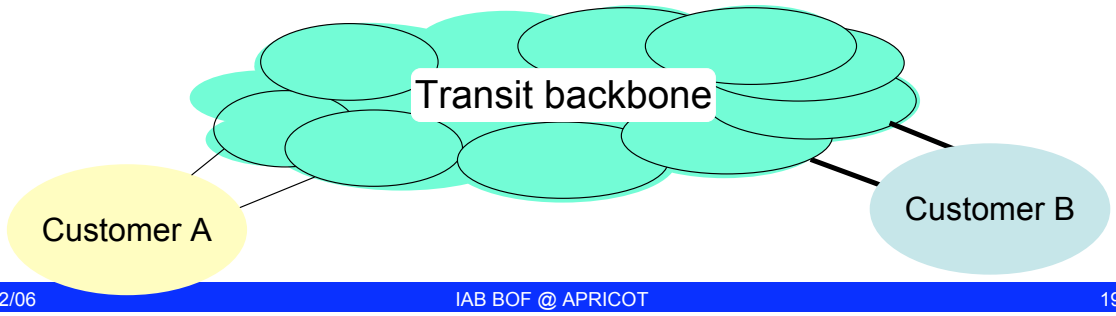
3. Arriving at destination

What Problems GSE Solved

- Making customer sites unaware of multihoming or provider change
  - Eliminate renumbering due to changing providers
- Providing freedom for whatever aggregation needed in the provider space

What Issues Left Open

1. *Inbound traffic engineering*: which destination RG to put on each packet?
2. *Outbound traffic engineering*: How to select exit router?
3. *Transit traffic engineering*: How to tell whether packets carrying different destination RGs belong/not belong to the same destination site?



What New Issues Introduced

- Interactions with DNS: given one finds destination RG from DNS lookup
  - Would the RGs for root servers (ever) change?
  - In general RGs for DNS servers *will* change, thorough analysis needed to understand the implication and impact
- What if remote link of the selected destination RG fails?
  - It is proposed not to use dynamic DNS for link status changes
- IP Tunnels that cross RG boundaries
  - What source/destination RGs going into the tunnel? How to handle the RGs when packets get out (land in a different site)?
- Concerns about whether allowing address rewriting would make TCP connection hijacking much easier

Multihoming: Summary of Issues

- PI prefix is effective for TE, but raises scaling concerns
- Use of multiple PA prefixes bring up new issues
  - Source address selection
  - Source traffic exit router selection
  - Destination address selection (affecting destination entry point)
  - Who/where/how to control the above
  - Transit traffic engineering

Other related issues not discussed

- Ingress filtering

A Few Departing Words

- GSE brought up a new approach to the problem
  - An interesting, but early stage, proposal; a number of important issues remain open
- (Quote from NANOG35 IAB BOF) "routing has always been working with prefixes, which are *locators*"
  - locator/identifier overload/split is an issue for transport and above, but *not* the cause of today's routing scalability problem
- (Tony Li) creating a scalable routing subsystem is paramount, as without that, we effectively have no network.
- The community must work together, *step by step*, towards a solution

Questions?

lixia@cs.ucla.edu