A How-To Guide to BGP Multihoming

Lane Patterson <lane@equinix.com>
Louis Lee <louie@equinix.com>
Equinix, Inc.

How To Use This Paper

This paper is geared toward readers who have a reasonable background in corporate and Internet networking, but are not experts in BGP or backbone Internet routing architectures. If you are evaluating BGP Multihoming, and need to justify and plan for it internally, Section 2 and 5 will be of interest. If you are ready to plan router configurations for BGP Multihoming, then Section 3 and 4 will be of interest.

Executive Summary

BGP Multihoming is a method that many medium and large enterprises and content providers use to attach their networks to two or more Internet Service Providers using Border Gateway Protocol version 4. Since there is a lack of up-to-date public information on BGP Multihoming, many organizations have either avoided it, or have implemented severely limited BGP Multihoming architectures, and have thus been unable to maximize the value of this technology.

This white paper provides an updated overview of BGP Multihoming, including prerequisites, basic implementation examples and router configurations, and advanced techniques and technologies for optimizing cost, performance, and flexibility of your multihomed Internet platform.

We begin with a review of multihoming prerequisites. We discuss the requirement to run BGP with your providers, and review the basics of this well-known routing protocol. We provide an overview of the process of registering an Autonomous System Number (ASN), which is a required identifier that allows BGP to group sets of routes by the organization that announces them. We review the way routes are propagated throughout the Internet by ISPs, and why this impacts your IP addressing plans. In particular, if you obtain public IP address space from an upstream ISP, you need a minimum of a /24 address block to implement BGP Multihoming. If your organization has a requirement for at least a /20 (Equivalent to sixteen /24's), you can register for address space directly from an addressing authority such as ARIN (www.arin.net), and enjoy the benefits of portable address space. Finally, we present the basic router performance requirements for running BGP—requirements that are increasing within the scope of standard enterprise routers.

Three complete BGP Multihoming examples are presented, including network diagrams, commentary, and router configurations. We consider the most common cases: single site, single router; and single site, dual routers. We show the difference between using provider-assigned address space and portable address space. We show how to verify BGP policy propagation by using public route views servers.

Advanced BGP techniques for static load balancing are presented. For outbound load balancing, we provide examples of preferencing routes based on large, well-known intermediate ASNs. For inbound load balancing, we present prefix engineering, AS-path prepending, and provider-supported BGP communities, as available tools. Despite all these techniques, we point out that standard BGP "tweaking" is not based on real time performance, traffic, or cost considerations, but simply manipulation of a routing protocol designed to ensure network connectivity.

We present some building blocks of advanced multihoming strategies and technologies, designed to provide better cost, performance and flexibility efficiencies. On the business side, we note the important role that ISP bandwidth contract negotiation can have, and the benefits of carrier-neutral colocation for eliminating local loops and providing a competitive bandwidth market. In such an environment, it is

possible to move from a scenario of "single-homed to a single, expensive provider" towards a scenario of "multihomed with majority of traffic going through an inexpensive competitive provider, and a low-commit burstible backup to a larger, more expensive, provider."

We present two new technologies that introduce cost and performance benefits. The first are multihoming platforms: services at carrier-neutral locations that automated the process of multihoming, and allow you to connect to multiple ISPs through a single router port[1]. The second technology is known as "route control." For sites that are multihomed, route control solutions seek to overcome the limitations of standard BGP by continually measuring the performance and cost of bandwidth through each provider link, and dynamically preferencing BGP routes based on performance and/or cost.[2]

In conclusion, we note that BGP Multihoming is continuing to evolve in ways that can help organizations meet their Internet performance, reliability, and redundancy goals, without increasing costs. We plan to continue to publish revisions of this white paper based on feedback and market developments, in order to promote the most effective use of BGP Multihoming.

1. Introduction

This white paper was developed to promote effective BGP Multihoming by medium and large enterprise and content sites. Despite the strategic significance of BGP Multihoming, it remains a poorly documented and often misunderstood topic for a number of reasons:

- ISPs do not promote multihoming because it reduces customer dependence on their services
- Many prior online multihoming tutorials are incomplete or have not been kept up to date as "living documents"
- Organizations that are prime candidates for multihoming may avoid it due to perception of cost, complexity, or technical expertise required.

This paper is geared to readers who have a reasonable background in corporate and Internet networking, but are not experts in BGP or core backbone routing architectures. It assumes you work for an organization with mission critical Internet requirements and sizable Internet traffic. Given today's realistic environment of cost control, we maintain a practical focus on how to use BGP Multihoming to do more with less, including bandwidth and local loop cost reduction strategies and new route control technologies. We have aimed to provide the kind of How-To information that will allow you to evaluate, budget for, propose and defend, design, implement, and optimize, a BGP Multihoming solution for your organization.

This paper is organized as follows:

- Section 2. Multihoming Prerequisites: This section presents items that need to be considered in preparation for BGP Multihoming: router performance requirements, IP address space, basic BGP and routing concepts, and obtaining an Autonomous System Number (ASN).
- Section 3. Complete Multihoming Examples: This section presents diagrams, router configurations (in Cisco IOS format), and commentary on the three most common BGP Multihoming configurations. It presents techniques and resources for verifying and troubleshooting a multihomed routing policy.

[1] For example, a product called Equinix Direct is an automated platform for multihoming through a single router port. Its strengths include no bandwidth commits, month-to-month terms, web-driven selection of providers by price and brand, and automated configuration when switching providers.
[2] For example, Equinix Intelligent Routing Service is a route control service based on an appliance from netVMG (www.netvmg.com). Sockeye and Route Science are examples of other route control solutions.

- Section 4. Advanced BGP Techniques--Static Load Balancing: This section discusses standard BGP configuration techniques and examples for load balancing inbound and outbound traffic across multiple provider links, and identifies the limitations of these techniques.
- Section 5. Multihoming Strategies and Technologies: This section discusses practical business and cost considerations when multihoming, such as bandwidth contract negotiation and optimization, ISP selection and switching, and elimination of local loops in carrier-neutral facilities. It presents some pitfalls of 95th percentile traffic billing that can lead to double-billing if not managed properly. It discusses new BGP Multihoming technologies such as Intelligent Routing and Equinix Direct that can help lower cost and improve performance.
- Section 6. Conclusions.

2. Multihoming Prerequisites

So you want to know what it takes to multihome? Let's start by making sure we are clear of the definition of multihoming, then we'll talk about what it takes to do BGP, IP addressing requirements and options for multihoming, and how to make sure your router is beefy enough for the task.

2.1 Definition of Multihoming

BGP Multihoming: maintaining links to multiple Internet providers (usually 2 or 3) and using BGP to send routes and to receive full routing tables from these providers.

2.2 BGP Requirement

BGP is the routing protocol required to connect to multiple Internet providers according to common multihoming practices[3]. Its purpose is for propagating routing information between Autonomous Systems—separately administered networks, whether they are large or small ISPs, or multihomed content providers or enterprises. BGP is an Exterior Gateway Protocol (EGP), and so is designed for the routing policy requirements that exist when two or more separate networking organizations interconnect. By contrast, an Interior Gateway Protocol, such as OSPF[4], is designed to propagate routes within a network that is under the control of a single organization, and does not have the level of sophistication and policy control to represent and control routing among multiple organizations.

Each Autonomous System must have a unique Autonomous System Number (ASN). Think of an ASN as a meta-label that is applied to the set of routes from an organization. Whether you are a small multihomed organization with one or two public Internet routes, or a large ISP with 25,000 Internet routes, BGP will tag all the routes from an AS with the ASN, to keep track of the network entities that must be traversed to reach a destination network.

BGP is not rocket science and is bundled in most routing software running on popular routers. But it does require some conscientious effort from a network engineer and an awareness of existing best practices and configuration guidelines, to set up a BGP routing session with a provider. BGP replaces the default route common in single-homed configurations with a protocol that dynamically informs your router of every specific routing prefix in use on the Internet (aggregated according to CIDR best practices), and the associated BGP attributes for each route.

BGP as it pertains to common multihoming examples is described in more detail in the section *"BGP Requirement: Overview and Concepts"* later in this document.

[3] Border Gateway Protocol version 4 (BGP) is an Internet Standard routing protocol described in RFC 1771 (http://www.ietf.org/rfc/rfc1771.txt).
[4] Open Shortest Path First version 2 (OSPFv2) is an Internet Standard routing protocol described in RFC 2328 (http://www.ietf.org/rfc/rfc2328.txt).

2.3 ASN Requirement: HOW TO for Getting an ASN

The first step in implementing multihoming is to obtain a publicly registered Autonomous System Number (ASN). In North America, the Caribbean and sub-Saharan Africa, ARIN (www.arin.net) is the organization delegated the responsibility of assigning ASNs[5]. Other countries and regions have their own registries for obtaining ASNs.[6]

The process of registering for an ASN from ARIN is described in detail at

http://www.arin.net/library/training/asn_process/index.html

While ARIN is the final authority for this process, here are some pointers and highlights:

1. Complete and submit the ARIN ASN Request Template. Make sure you take the time to review this carefully, and if you are unsure of anything, get outside review from consultants or sources that are knowledgeable about the process.
2. ARIN should give you acknowledgement and a tracking number for you submission, and respond within 3 working days.
3. To save time, start your review and preparation of the ARIN Registration Services Agreement and the ASN Billing Account Form. Once your ASN Request Template is approved, ARIN will ask you to submit these.
4. The Credit Card option is the fastest form of payment.
5. Once ARIN has received and processed everything, they will issue your ASN. You should be able to query the ARIN database for your ASN and point of contact (POC) information within a couple days of receiving notice from ARIN. Take time to verify this at http://www.arin.net/tools/whois_help.html
6. You will probably find it useful to register your maintainer ID, ASN, and routing announcements with a routing registry. In North America, RADB is commonly used (http://www.radb.net/about.html) Many providers use the RADB to determine valid routing announcements, and may not listen to your routes unless they match the policy you register in the RADB.

2.4 IP Addressing/Routing Requirement

Unless you have sufficient IP addressing requirements, you probably have received your IP subnets (a.k.a. routes, prefixes, netblocks) from an upstream ISP. These subnets are commonly part of a larger block of address space that your ISP has been assigned by a registry such as ARIN. This type of IP address prefix is known as PA-space, for *Provider-Assigned*.

Organizations within the ARIN territory that have demonstrated a requirement for more than a /21 (Eight /24's) can request a minimum of a /20 (Sixteen /24's) of IP address space directly from ARIN, according to their guidelines at http://www.arin.net/policy/ipv4.html. This type of IP address space is known as PI-space, for *Provider-Independent*. Just as with the process for requesting an ASN, be prepared to familiarize yourself with ARIN's process for documenting IP address requirements and correctly reviewing and submitting applications. If your documentation is clean, you will find that it is a smooth process.

The minimum prefix you must have to multihome is a /24. If your current provider-assigned subnet is smaller, such as a /27, then you will need to work with your provider to request a full /24 assignment, justified by a request to run BGP and do multihoming. Since this may involve re-numbering, you will want to pick the provider you use for your /24 assignment carefully. Regardless of which other providers to which you multihome, you will necessarily have to retain the provider that supplies your address space.

[5] ARIN's policies for registering an Autonomous System Number are published at www.arin.net/policy/asn.html

[6] Links to registries around the world can be found at www.arin.net/library/internet_info/countries.html

You also need to consider the "routability" of your address blocks, something that ARIN and other registries don't guarantee (they just assign the space—routing it is up to you and your providers). *Routability* simply means that you announce your prefix(es) via BGP to your upstream providers, and that they in turn announce them to their peers so that they are visible throughout the Internet. However, you must always keep in mind that the most specific route always wins. For example, if your primarily provider only announces your /24 as part of a /13 prefix, and your secondary provider announces it as a specific /24 prefix, then all your inbound traffic will come in through your secondary provider. If this is important to you, get agreement from your providers on how they will propagate your route announcements, and whether any major peers are likely to ignore them.

To re-cap IP addressing and routability issues with Provider-Allocated (PA) space:

1. Make sure your existing provider supports your requirement for a minimum of /24 for the purpose of BGP Multihoming
2. Determine fees and policies and procedures for the address space and BGP setup with each provider
3. Ensure that your existing provider is willing to propagate your BGP announcement of your discrete address block (e.g. /24) to their peers, in addition to the entire aggregated prefix (e.g. /13) that they would normally announce. This need not apply if you have no requirement to load balance inbound traffic, and don't mind if your inbound traffic comes in primarily through your secondary provider(s).
4. Ensure that your secondary provider(s) support your requirement to run BGP with them to propagate your prefix(es) to their peers. (*Note: their peers are not necessarily obliged to accept these more-specific announcements, and may filter such that they only listen to the aggregated prefix announced by your primary provider. This is not a widespread practice, but is known to be followed by two large ISPs)

If you are able to justify Provider-Independent (PI) space from your address registry, you will not be anchored to one "primary" ISP, and can freely switch ISPs without renumbering your addresses. You will not need to worry about the issues above that apply to PA-space.

2.5 Router Hardware, Memory, and CPU Requirements

The final requirement for multihoming is router horsepower. For starters, you'll need extra ports on your router for multiple ISP connections, sized according to your current and planned traffic requirements.

The other main requirement is for RAM. Receiving full BGP routing tables from your providers requires at least 128 MB RAM, and preferably 256 MB RAM. RAM requirements depend on how many transit providers you will have (this is a variable you control), as well as how much natural growth occurs in the BGP routing tables (this is a variable you cannot control, and it pays to allow for growth room of 30-50%). You will also want to keep a minimum of at least 30 MB free RAM. Of course RAM requirement is also a function of any other features you choose to run on your router.

Here's one example of RAM utilization: In early 2003, a Cisco 7206 VXR with 256 MB RAM is configured with minimal feature use and full BGP tables from 4 different providers. The size of the BGP tables averages 118,000 routes. RAM utilization is observed to be about 108MB. Router resources won't be an issue for a long time.

A third requirement to keep in mind is CPU utilization. Running BGP, and related processes like BGP route flap dampening (a common feature to reduce instability in your network caused by unstable routes elsewhere on the Internet), take a certain amount of CPU processing power. On properly sized routers, this is usually not a problem. But don't cut it too close. I saw one example of a company multihomed to 3 providers using a Cisco 3640 with 128MB RAM. Since the Cisco 3640 doesn't have hardware-based forwarding to the extent that other router makes/models do, CPU utilization was also influenced by their daily traffic peaks. During traffic of 20-30 mbps, the router was comfortable at 35% CPU utilization. But when traffic occasionally peaked at 40-60 mbps, the CPU became saturated, and an immediate router upgrade was required. By contrast, routers with fully distributed hardware-based forwarding usually see very low CPU loads, even when running BGP to multiple upstream providers and pushing traffic at line rate.

This brings us to the topic of traffic throughput. Keep in mind that just because a router can be configured with a FastE or GigE interface, doesn't mean that the router has the horsepower to fill up that interface. For instance, a Cisco 7200 uses PCI busses internally for packet forwarding, and you should generally limit total traffic throughput to less than 300 mbps. Extreme and Foundry switches, while excellent at forwarding at line rate as a layer 2 ethernet switch, are generally much more CPU-intensive when used as IP routers. Be careful to do your research on what to expect when you combine your BGP, IP routing, traffic volume, and feature requirements on the same box, so you don't get surprised. At the high end, Juniper has an excellent performance to price ratio on the M5 and M10 platforms, and has a proven reputation for being able to run full line rate with important features enabled. Cisco also has newer, more robust, routers such as the 7300 family, with PXF hardware acceleration and 2 built-in GigE ports. If you are handling more than 200-300 mbps, this platform may give you room for growth without requiring a costly GSR purchase. The 7600 series is also generally a good value for high-end access needs, and provides a single box to combine the switching and routing requirements of many end sites. This is by no means a complete analysis, and it won't take long for this information to become obsolete, so make sure to get current recommendations at the time you develop your multihoming requirements.

2.6 BGP Requirement: Overview and Concepts

So you are ready to run BGP! This section is not meant to be an exhaustive BGP tutorial—there are plenty of more in-depth resources available for that. What it does cover is the high level concepts that may be new to you if you are used to a single-homed setup.

Default Routing: This is an easy way to route traffic when there is only one path available for all destinations. You simply configure the same next hop router for ANY destination. For example, hosts such as desktop and server PCs are usually configured with a *default gateway*.

Full Routes: This is the term for receiving every routable Internet prefix explicitly, so that you do not need a default route to your Internet providers. As of early 2003, there are approximately 120,000 routes in a full BGP table for Internet transit. Depending on the routing policies of a given ISP, this number can vary between approximately 105,000 routes and 135,000 routes. Given that there are between 400-800 million Internet users world wide, routes typically represent highly aggregated sets of end users. Default routing does not work with multihoming, since the whole point is to pick between multiple possible links for every Internet destination. Receiving full routes gives your router the information needed to effectively multihome.

Asymmetric Routing: Internet routing is a not a symmetric proposition. In order for you to receive inbound traffic, you or your providers must announce the route to get to you. In order for you to respond with outbound traffic, you must receive and accept routes back to others. The two processes are completely decoupled and independent. Whereas most single-homed customers are used to thinking of all their inbound and outbound traffic going through the same path, the complete traceroute inbound versus outbound is almost always different. *Asymmetric routing is the norm in the core of the Internet.*

Announcing Routes: *Outbound route announcements influence inbound traffic.* Running BGP means that your route announcements are now dynamic, not static. It is up to you to announce your prefixes to your upstream providers, and for them to pass these prefixes on to other networks. The distribution of your routing information throughout the Internet allows end networks to learn how to get to you. This is a benefit, for instance, when your link to one ISP fails, because BGP will automatically withdraw route announcements through that path, and your inbound traffic will fail over through the remaining path(s).

There are specific BGP techniques available (discussed more in Section 4.) that can help you influence which inbound path will be used to reach your site. But remember that ultimately, these techniques only pass on "hints" to other networks, and the final decision how to route inbound to you is really up to them.

Best Practices for Announcing Routes: The most important best practice for announcing routes is to make sure you apply a prefix-filter to the routes you announce to your providers that restricts the announcements to your specific aggregated netblocks. This prevents two important things that would otherwise occur with BGP: (1) announcing routes you learned from “provider A” back out to “provider B” (which would inadvertently make you an alternative path between provider A and B!) and (2) announcing both your aggregate prefix, and all those nifty internal subnets you’ve broken it into for internal purposes.

Accepting Routes: *Routes received and accepted influence outbound traffic.* When each of your transit providers is announcing valid routes to all destinations on the Internet, you have the ability to make choices about which provider to choose, which affects your outbound traffic load balancing.

Unlike routing of inbound traffic, you are in complete control of the next-hop selection of outbound traffic. How you set preferences on the routes you receive will determine which outbound transit provider is preferred for a given destination. There are a number of ways to set preferences on received routes. BGP has a standard selection process for determining which route is preferred, when several are available to the same destination prefix[7]. This decision process allows plenty of hooks for you to set your own choices, commonly via the BGP local preference attribute. For example, you select subsets of received routes based on BGP communities (e.g. all routes with the “customer community” value set by your upstream provider), AS-path operations (e.g. all routes destined to AOL’s ASN), or even specific prefixes. Once you match these subsets of routes, you can perform an action on the subset, such as setting local-preference higher or lower, to help prefer or avoid the given path.

Best Practices for Accepting Routes (from Transit Providers): We must distinguish the practices for accepting routes from transit providers versus peering partners, as they are very different. The most important things to filter out when accepting routes from transit providers are (1) “bogon” networks, such as RFC1918 private networks (e.g. 10.0.0.0/8) that you may use internally, and which should never be routed across the public Internet, and (2) limiting the maximum number of routes you receive from your transit provider, so a mistake on their part does not starve the memory on your router. (e.g. set max-prefix to 140,000 if you commonly receive 120,000 routes from your provider. Update these numbers every few months.)

Other operations you will commonly perform on accepting routes and setting routing preferences fall into the “performance tuning” or “traffic engineering” category, and are not mandatory best practice.

Many good best practices with router configuration examples and explanations are described at http://www.cymru.com/Documents/secure-bgp-template.html.

iBGP mesh requirement: When you setup BGP to an external ASN, such as your upstream providers, it is known as External BGP (eBGP). If you have multiple Internet routers in your own network, you also have a requirement to run Internal BGP (iBGP) between them. iBGP sessions are generally more simple to setup, but they are absolutely required. All routers that will be routing from your ASN to the public Internet must have a full mesh of iBGP sessions among them. For a pair of Internet routers, this is simply one iBGP session. But for 4 Internet routers, this adds up to 6 iBGP sessions to complete the full iBGP mesh.

We have presented important high-level BGP concepts, without getting into specific mechanisms for accomplishing them. In the next section, we will show some complete examples of common multihoming cases, including example router configurations for addressing and routing.

[7] The BGP Decision Algorithm is presented in http://www.nanog.org/mtg-9901/ppt/bgp102/sld020.htm

3. Complete Multihoming Examples

This section covers the most common basic multihoming scenarios. Each scenario consists of a network diagram, description of the scenario, and an example of the BGP router configuration required to implement it. Cisco IOS configuration commands are used. Note that these are "bare bones" configs—no advanced load balancing or preferencing route maps are applied. These will be described in separate "config snippets" in the Advanced BGP Techniques section.

3.1 Example 1: Single Site, Single Router, PA-space, Multihoming to Two ISPs

Background: In this example, a single organization is connecting at a single site, using a single router, to two ISPs (without regard to which one is primary and/or secondary). It is receiving full transit BGP routes from each provider, and is allocated Provider-Assigned (PA) address space from the first provider, ISP 1. Figure 1 shows the network architecture and relevant details for this example.

Note: When PA-space is used, the provider that owns the address block (in this case, ISP 1) will announce their aggregate block to the rest of the Internet, as well as (if you arrange it properly with them) propagating your specific /24 route announcement originated from your ASN. The caveat with this case is that other Internet peers are not required to accept the more specific route. While this is not common, it may occur that a 3rd party ISP could only reach you through the ISP1 path, based on the aggregate route announcement. If your link to ISP1 goes down, these 3rd party ISPs could end up still trying to route through ISP1, unaware that a more specific route exists through ISP2.

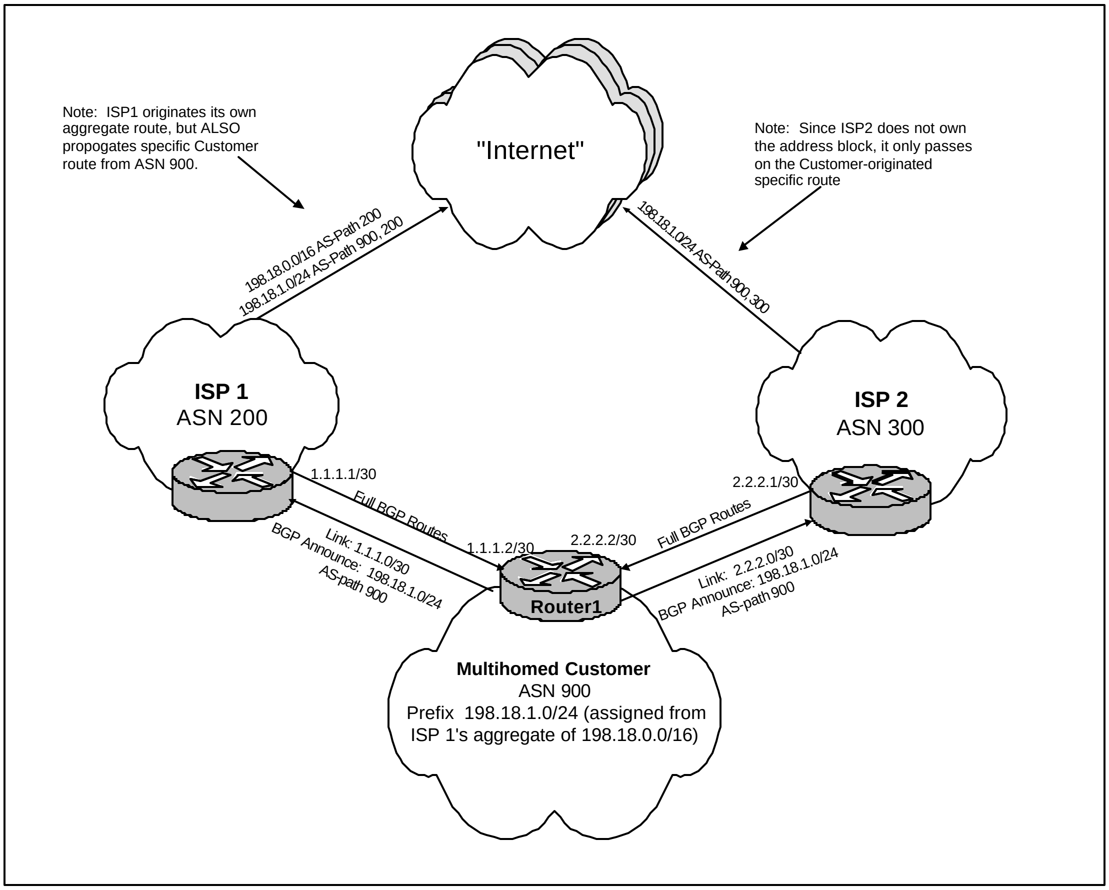

Note: ISP1 originates its own aggregate route, but ALSO propogates specific Customer route from ASN 900.
"Internet"
Note: Since ISP2 does not own the address block, it only passes on the Customer-originated specific route
198.18.0.0/16 AS-Path 200
198.18.1.0/24 AS-Path 900, 200
198.18.1.0/24 AS-Path 900, 300
ISP 1
ASN 200
ISP 2
ASN 300
1.1.1.1/30
Full BGP Routes
2.2.2.1/30
Full BGP Routes
1.1.1.2/30
2.2.2.2/30
Link: 1.1.1.0/30
BGP Announce: 198.18.1.0/24
AS-path 900
Link: 2.2.2.0/30
BGP Announce: 198.18.1.0/24
AS-path 900
Router1
Multihomed Customer
ASN 900
Prefix 198.18.1.0/24 (assigned from ISP 1's aggregate of 198.18.0.0/16)

Figure 1. Multihoming: PA-space Addresses, Single Site, Single Router, Two ISPs

Example 1 Router Configuration

```
! Single-router, PA-space, Basic Config
!
! Define your BGP ASN on your router
autonomous-system 900
!
! Define static NULL route to networks that
! will be announced to providers via BGP.
ip route 198.18.1.0 255.255.255.0 Null0 200
!
! Define ANNOUNCE prefix list, of your netblocks to announce
! via BGP to your providers.  You will apply this prefix-
! list outbound on the BGP session to each provider.
ip prefix-list ANNOUNCE description Our External Netblocks
ip prefix-list ANNOUNCE seq 10 permit 198.18.1.0/24
!
! Define BOGONS prefix list, of bad netblocks you
! need to block from being accepting from your providers.
! Don’t just trust your provider to run a clean network!
! You will apply this prefix-list inbound on the BGP
! session to each provider.
ip prefix-list BOGONS description Bad Routes to Block In
ip prefix-list BOGONS seq 10 deny 0.0.0.0/8 le 32
ip prefix-list BOGONS seq 15 deny 10.0.0.0/8 le 32
ip prefix-list BOGONS seq 20 deny 127.0.0.0/8 le 32
ip prefix-list BOGONS seq 25 deny 172.16.0.0/12 le 32
ip prefix-list BOGONS seq 30 deny 192.0.2.0/24 le 32
ip prefix-list BOGONS seq 35 deny 192.168.0.0/16 le 32
ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32
! Prevent someone else from announcing your own prefix(es)
! back to you, for security: update this with YOUR
! actual announced block(s)!
ip prefix-list BOGONS seq 1000 deny 198.18.1.0/24 le 32
! Accept any other routes bigger or equal to /27.  You
! can tweak this up to /24 if you like.
ip prefix-list BOGONS seq 9999 permit 0.0.0.0/0 le 27
!

!
router bgp 900
 ! don’t require your IGP to be in synch with BGP,
 ! synchronization has been outmoded for some time.
 no synchronization
 ! tell your  router to log changes to your BGP
 ! sessions, you’ll want to be concerned with BGP
 ! sessions when they go up and down, it’s just as
 ! important to your routing as a link up/down.
 bgp log-neighbor-changes
 ! enable BGP dampening to minimize adverse impact
 ! of “flapping” routes (routes that are announced
 ! and withdrawn repeatedly).
 bgp dampening
 ! define your BGP network statements: these are the
 ! aggregate external IP blocks you will be announcing
 ! to the Internet.  Note that the network statement
 ! will not be effective unless there is an underlying
 ! route for the network, which is why we defined a
 ! static NULL route for this block above.
 network 198.18.1.0 mask 255.255.255.0
 ! define our BGP session with ISP-1 (ASN 200)
 !
 neighbor 1.1.1.1 remote-as 200
 ! description allows you to put add a text label
 neighbor 1.1.1.1 description BGP Transit to ISP-1
 ! hard-code version 4 to short-cut BGP version negotiation
 neighbor 1.1.1.1 version 4
 ! send-community is nice if you will be setting communities
 ! on routes you announce to influence how your upstream
 ! provider re-announces the routes to the Internet.  Many
 ! providers support sophisticated community sets to allow
 ! this kind of customer control.
 neighbor 1.1.1.1 send-community
 ! Soft reconfiguration is nice, it prevents complete
 ! withdrawal and relearning of routes when doing “clear
 ! ip bgp” command.  But it does require enough RAM to
 ! cache an extra copy of the table.
 neighbor 1.1.1.1 soft-reconfiguration inbound
 ! Filter out bogus prefixes from your upstream.  Don’t
 ! trust your ISP to do this for you.
 neighbor 1.1.1.1 prefix-list BOGONS in
 ! Limit your announcement just to your public prefix(es).
 ! This enforces aggregation, and prevents you from
 ! announcing ASN 200’s routes to ASN 300, which would
 ! accidentally make yourself a transit between the two
 ! ISPs.
 neighbor 1.1.1.1 prefix-list ANNOUNCE out
 ! enforce max-prefix limit: just in case your provider
 ! blows up their routing tables, this keeps your router
 ! from melting under the stress by shutting off the
 ! mis-behaving BGP session instead.  Once your ISP fixes
 ! the problem, you can re-enable with a “clear ip bgp ...”
 neighbor 1.1.1.1 maximum-prefix 140000

 ! define BGP session with ISP-2 (ASN 300)
 neighbor 2.2.2.1 remote-as 300
 neighbor 2.2.2.1 description BGP Transit to ISP-2
 neighbor 2.2.2.1 version 4
 neighbor 2.2.2.1 send-community
 neighbor 2.2.2.1 soft-reconfiguration inbound
 neighbor 2.2.2.1 prefix-list BOGONS in
 neighbor 2.2.2.1 prefix-list ANNOUNCE out
 neighbor 2.2.2.1 maximum-prefix 140000
 !
end
```

3.2 Example 2: Single Site, Dual Routers, PA-space, Multihoming to Two ISPs

Background: This example is the same as Example 1, except the customer is using two routers, and connecting to one ISP on each router. The purpose of this example is to show how iBGP is used to internally propagate BGP routes between the two routers.

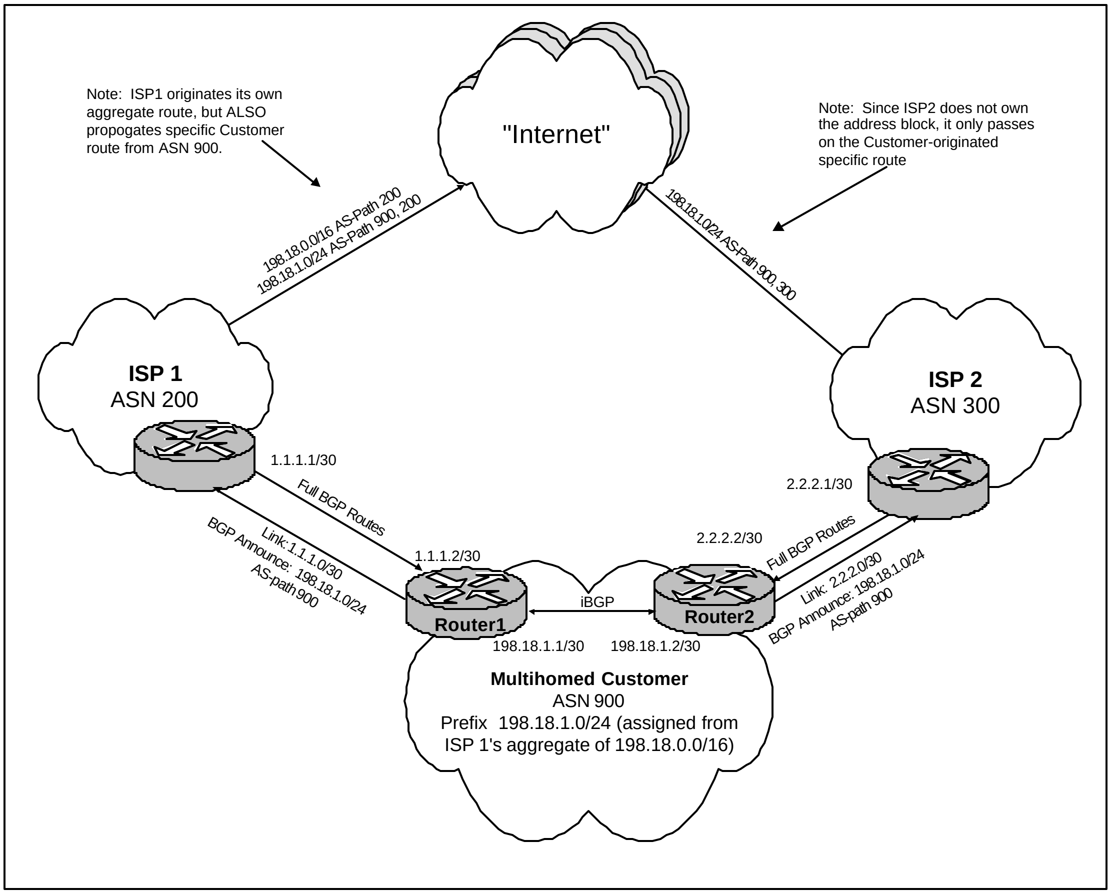

Note: ISP1 originates its own aggregate route, but ALSO propogates specific Customer route from ASN 900.
"Internet"
Note: Since ISP2 does not own the address block, it only passes on the Customer-originated specific route
198.18.0.0/16 AS-Path 200
198.18.1.0/24 AS-Path 900, 200
198.18.1.0/24 AS-Path 900, 300
ISP 1
ASN 200
ISP 2
ASN 300
1.1.1.1/30
Full BGP Routes
Link: 1.1.1.0/30
BGP Announce: 198.18.1.0/24
AS-path 900
1.1.1.2/30
2.2.2.1/30
2.2.2.2/30
Full BGP Routes
Link: 2.2.2.0/30
BGP Announce: 198.18.1.0/24
AS-path 900
Router1
iBGP
Router2
198.18.1.1/30
198.18.1.2/30
Multihomed Customer
ASN 900
Prefix 198.18.1.0/24 (assigned from ISP 1's aggregate of 198.18.0.0/16)

Figure 2. Multihoming: PA-space Addresses, Single Site, Dual Routers, Two ISPs

Example 2 Router Configurations

```
!
! Router 1:  Dual-router, PA-space, Basic Config
!
autonomous-system 900
!
ip route 198.18.1.0 255.255.255.0 Null0 200
!
ip prefix-list ANNOUNCE description Our External Netblocks
ip prefix-list ANNOUNCE seq 10 permit 198.18.1.0/24
!
ip prefix-list BOGONS description Bad Routes to Block In
ip prefix-list BOGONS seq 10 deny 0.0.0.0/8 le 32
ip prefix-list BOGONS seq 15 deny 10.0.0.0/8 le 32
ip prefix-list BOGONS seq 20 deny 127.0.0.0/8 le 32
ip prefix-list BOGONS seq 25 deny 172.16.0.0/12 le 32
ip prefix-list BOGONS seq 30 deny 192.0.2.0/24 le 32
ip prefix-list BOGONS seq 35 deny 192.168.0.0/16 le 32
ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32
! Prevent someone else from announcing your own prefix(es)
! back to you, for security: update this with YOUR
! actual announced block(s)!
ip prefix-list BOGONS seq 1000 deny 198.18.1.0/24 le 32
ip prefix-list BOGONS seq 9999 permit 0.0.0.0/0 le 27
!
router bgp 900
 no synchronization
 bgp log-neighbor-changes
 bgp dampening
 network 198.18.1.0 mask 255.255.255.0
 neighbor 1.1.1.1 remote-as 200
 neighbor 1.1.1.1 description BGP Transit to ISP-1
 neighbor 1.1.1.1 version 4
 neighbor 1.1.1.1 send-community
 neighbor 1.1.1.1 soft-reconfiguration inbound
 neighbor 1.1.1.1 prefix-list BOGONS in
 neighbor 1.1.1.1 prefix-list ANNOUNCE out
 neighbor 1.1.1.1 maximum-prefix 140000

 ! define iBGP session
 neighbor 198.18.1.2 remote-as 900
 neighbor 198.18.1.2 description iBGP to Router2
 neighbor 198.18.1.2 version 4
 neighbor 198.18.1.2 send-community
 neighbor 198.18.1.2 soft-reconfiguration inbound
 ! make sure you are using a Loopback0 /32 on each
 ! router, and propagating this route internally.
 ! This is important to the stability of your iBGP
 ! sessions, so they are not tied to a physical
 ! interface address.
 neighbor 198.18.1.2 update-source Loopback0
!
end
```

```
!
! Router 2:  Dual-router, PA-space, Basic Config
!
autonomous-system 900
!
ip route 198.18.1.0 255.255.255.0 Null0 200
!
ip prefix-list ANNOUNCE description Our External Netblocks
ip prefix-list ANNOUNCE seq 10 permit 198.18.1.0/24
!
ip prefix-list BOGONS description Bad Routes to Block In
ip prefix-list BOGONS seq 10 deny 0.0.0.0/8 le 32
ip prefix-list BOGONS seq 15 deny 10.0.0.0/8 le 32
ip prefix-list BOGONS seq 20 deny 127.0.0.0/8 le 32
ip prefix-list BOGONS seq 25 deny 172.16.0.0/12 le 32
ip prefix-list BOGONS seq 30 deny 192.0.2.0/24 le 32
ip prefix-list BOGONS seq 35 deny 192.168.0.0/16 le 32
ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32
! Prevent someone else from announcing your own prefix(es)
! back to you, for security: update this with YOUR
! actual announced block(s)!
ip prefix-list BOGONS seq 1000 deny 198.18.1.0/24 le 32
ip prefix-list BOGONS seq 9999 permit 0.0.0.0/0 le 27
!
router bgp 900
 no synchronization
 bgp log-neighbor-changes
 bgp dampening
 network 198.18.1.0 mask 255.255.255.0
! define BGP session with ISP-2 (ASN 300)
 neighbor 2.2.2.1 remote-as 300
 neighbor 2.2.2.1 description BGP Transit to ISP-2
 neighbor 2.2.2.1 version 4
 neighbor 2.2.2.1 send-community
 neighbor 2.2.2.1 soft-reconfiguration inbound
 neighbor 2.2.2.1 prefix-list BOGONS in
 neighbor 2.2.2.1 prefix-list ANNOUNCE out
 neighbor 2.2.2.1 maximum-prefix 140000

 ! define iBGP session
 neighbor 198.18.1.1 remote-as 900
 neighbor 198.18.1.1 description iBGP to Router1
 neighbor 198.18.1.1 version 4
 neighbor 198.18.1.1 send-community
 neighbor 198.18.1.1 soft-reconfiguration inbound
 ! make sure you are using a Loopback0 /32 on each
 ! router, and propagating this route internally.
 ! This is important to the stability of your iBGP
 ! sessions, so they are not tied to a physical
 ! interface address.
 neighbor 198.18.1.1 update-source Loopback0
!
end
```

3.3 Example 3: Single Site, Dual Router, PI-space, Multihoming to Two ISPs

Background: This example is the same as Example 2, except the customer has qualified for and is using an ARIN-allocated /19 address block, instead of a provider-assigned /24. Figure 3 shows the difference in how this routing announcement propagates outward throughout the Internet, since it is not covered by an overlapping aggregate announcement by ISP 1.

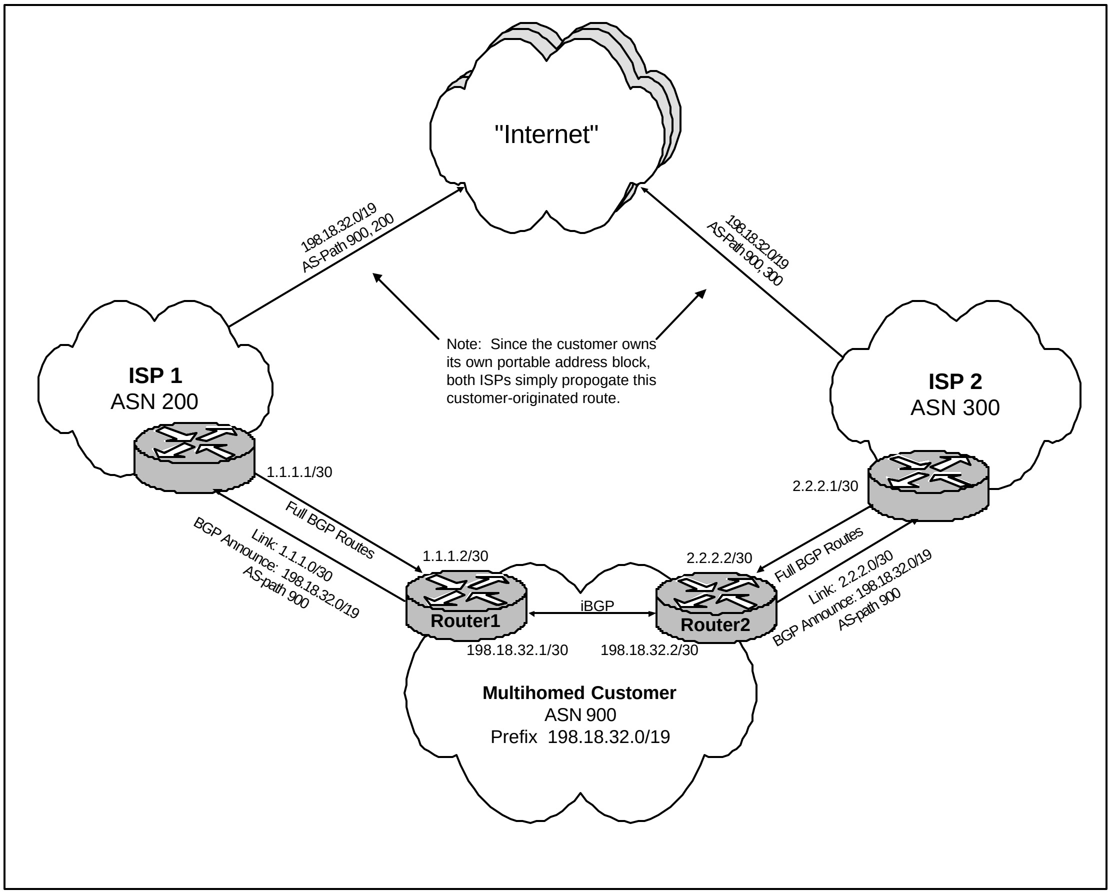


**Figure 3. Multihoming: PI-space Addresses,
Single Site, Dual Routers, Two ISPs**

Example 3 Router Configuration

```
!
! Router 1:  Dual-router, PI-space, Basic Config
!
autonomous-system 900
!
! static Null0 route for 198.18.32.0/19
ip route 198.18.32.0 255.255.224.0 Null0 200
!
ip prefix-list ANNOUNCE description Our External Netblocks
ip prefix-list ANNOUNCE seq 10 permit 198.18.32.0/19
!
ip prefix-list BOGONS description Bad Routes to Block In
ip prefix-list BOGONS seq 10 deny 0.0.0.0/8 le 32
ip prefix-list BOGONS seq 15 deny 10.0.0.0/8 le 32
ip prefix-list BOGONS seq 20 deny 127.0.0.0/8 le 32
ip prefix-list BOGONS seq 25 deny 172.16.0.0/12 le 32
ip prefix-list BOGONS seq 30 deny 192.0.2.0/24 le 32
ip prefix-list BOGONS seq 35 deny 192.168.0.0/16 le 32
ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32
! Prevent someone else from announcing your own prefix(es)
! back to you, for security: update this with YOUR
! actual announced block(s)!
ip prefix-list BOGONS seq 1000 deny 198.18.32.0/19 le 32
ip prefix-list BOGONS seq 9999 permit 0.0.0.0/0 le 27
!
router bgp 900
 no synchronization
 bgp log-neighbor-changes
 bgp dampening
 network 198.18.32.0 mask 255.255.224.0
 neighbor 1.1.1.1 remote-as 200
 neighbor 1.1.1.1 description BGP Transit to ISP-1
 neighbor 1.1.1.1 version 4
 neighbor 1.1.1.1 send-community
 neighbor 1.1.1.1 soft-reconfiguration inbound
 neighbor 1.1.1.1 prefix-list BOGONS in
 neighbor 1.1.1.1 prefix-list ANNOUNCE out
 neighbor 1.1.1.1 maximum-prefix 140000

 ! define iBGP session
 neighbor 198.18.32.2 remote-as 900
 neighbor 198.18.32.2 description iBGP to Router2
 neighbor 198.18.32.2 version 4
 neighbor 198.18.32.2 send-community
 neighbor 198.18.32.2 soft-reconfiguration inbound
 ! make sure you are using a Loopback0 /32 on each
 ! router, and propagating this route internally.
 ! This is important to the stability of your iBGP
 ! sessions, so they are not tied to a physical
 ! interface address.
 neighbor 198.18.32.2 update-source Loopback0
!
end


!
! Router 2:  Dual-router, PA-space, Basic Config
!
autonomous-system 900
!
! static Null0 route for 198.18.32.0/19
ip route 198.18.32.0 255.255.224.0 Null0 200
!
ip prefix-list ANNOUNCE description Our External Netblocks
ip prefix-list ANNOUNCE seq 10 permit 198.18.32.0/19
!
ip prefix-list BOGONS description Bad Routes to Block In
ip prefix-list BOGONS seq 10 deny 0.0.0.0/8 le 32
ip prefix-list BOGONS seq 15 deny 10.0.0.0/8 le 32
ip prefix-list BOGONS seq 20 deny 127.0.0.0/8 le 32
ip prefix-list BOGONS seq 25 deny 172.16.0.0/12 le 32
ip prefix-list BOGONS seq 30 deny 192.0.2.0/24 le 32
ip prefix-list BOGONS seq 35 deny 192.168.0.0/16 le 32
ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32
! Prevent someone else from announcing your own prefix(es)
! back to you, for security: update this with YOUR
! actual announced block(s)!
ip prefix-list BOGONS seq 1000 deny 198.18.32.0/19 le 32
ip prefix-list BOGONS seq 9999 permit 0.0.0.0/0 le 27
!
```

```
router bgp 900
 no synchronization
 bgp log-neighbor-changes
 bgp dampening
 network 198.18.32.0 mask 255.255.224.0
! define BGP session with ISP-2 (ASN 300)
 neighbor 2.2.2.1 remote-as 300
 neighbor 2.2.2.1 description BGP Transit to ISP-2
 neighbor 2.2.2.1 version 4
 neighbor 2.2.2.1 send-community
 neighbor 2.2.2.1 soft-reconfiguration inbound
 neighbor 2.2.2.1 prefix-list BOGONS in
 neighbor 2.2.2.1 prefix-list ANNOUNCE out
 neighbor 2.2.2.1 maximum-prefix 140000

 ! define iBGP session
 neighbor 198.18.32.1 remote-as 900
 neighbor 198.18.32.1 description iBGP to Router1
 neighbor 198.18.32.1 version 4
 neighbor 198.18.32.1 send-community
 neighbor 198.18.32.1 soft-reconfiguration inbound
 ! make sure you are using a Loopback0 /32 on each
 ! router, and propogating this route internally.
 ! This is important to the stability of your iBGP
 ! sessions, so they are not tied to a physical
 ! interface address.
 neighbor 198.18.32.1 update-source Loopback0
!
end
```

3.4 Route Announcement Verification and Troubleshooting

Congratulations, if you've made it this far, you've successfully gone through the basics of setting up BGP Multihoming! Let's look at some tools that can help you determine if your BGP announcements are being correctly received by (1) your direct providers (from you), and (2) other providers in the Internet (from peering sessions with your providers).

Many providers maintain either a web-based "Looking Glass" site or a BGP router that is publicly available for telnet, so that customers and ISPs can remotely verify BGP announcements using commands like "`show ip bgp x.x.x.x`". After all, the only way for you to know if your routing announcements are "making it out across the Internet" is for you to be able to remotely check BGP tables in other providers' networks. A good listing of these resources can be found at any of the following links:

- http://www.traceroute.org
- http://www.traceroutes.com
- http://www.cymru.com/Documents/secure-bgp-template.html

For you to properly verify that your routing announcements are being propagated via BGP out to the Internet, you should:

1. Ask your direct providers for access to Looking Glass URLs or public route views that are available. Verify what your prefix announcements look like within your provider using "show ip bgp x.x.x.x". Look for both specific announcements (and verify from the AS-path that they originate from your ASN) as well as aggregate prefix blocks that are announced from you provider's ASN, in the case of PA-space. Do this for all of your transit providers.
2. Pick at least 2 or 3 other major providers in the Internet, and use their Looking Glass resources to verify that your routes are being correctly passed on to them. Determine if these providers are receiving all the available paths (e.g. one from each of your direct providers).

3.5 Real World Examples of Organizations that Multihome

Using Looking Glass tools, and public WHOIS servers from ARIN, it is possible to see other enterprises and organizations that multihome. Here are a few examples, at the time of this writing, based strictly on this type of public information.

Name (ASN)	Providers Used to Multihome
Alibris (ASN 20198)	UUnet and PBIS
CollegeBoard (ASN 16919)	UUnet and InterNAP
Costco (ASN 11283)	Single homed to InterNAP (Here's an example of a large enterprise who chose to forego multihoming, and chose a single provider who indirectly provides some of these benefits.)
Ebay (ASN 11643)	XO, C&W
EDS (ASN 2165)	UUnet, Sprint
FedEx (ASN 7726)	Sprint and AT&T
Juniper Networks (ASN 14203)	UUnet and C&W
International Paper (ASN 16988)	Sprint and AT&T
Paypal (ASN 17012)	UUnet and InterNAP
Schwab (ASN 6949)	AT&T, InterNAP, and UUnet
Wells Fargo (ASN 10837)	AT&T and Qwest

3.6 More Complex Multihoming Examples

We won't cover more complex examples of multihoming in this white paper, but we will mention some:

- *Diverse Sites, Same Metro Area, Interconnected By MAN Link:* example—a mid-sized or large corporation with major on-line presence has a hot backup data center 30 miles away from its headquarters, on the other side of town. They wish to multihome to a different ISP at each location, and use the MAN link as a backup if either ISP goes down.
- *Diverse Sites, Long Distance (e.g. East and West Coast), Interconnected By WAN Link:* example—the online e-Commerce portal of a diversified conglomerate has some of its brands served out of its San Jose data center, and the rest served out of its Chicago business unit. The company already has good WAN capacity between the two sites for database backup purposes, and wishes to leverage this for Internet transit redundancy as well.
- *Diverse Sites, Long Distance, Interconnected By Internet Tunnel or VPN Link:* example—a medium sized company in Los Angeles acquires another entity in New York. They wish to consolidate internet presence without additional WAN costs.
- *Augmenting Your IP Transit With Settlement-Based or Free Peering:* example—an online gaming company experiences high growth in bandwidth requirements. But due to its popularity with broadband users, it is able to strike settlement-free peering deals at a neutral exchange point to route directly to customers of major DSL and Broadband Cable consumer companies. This saves it 40% in bandwidth costs in year 1.

We also do not discuss some of the alternatives to doing BGP Multihoming:

- Multi-addressing: using a different subnet and static route for 2 or more providers, and using multiple addresses on web servers or load balancers to use these providers simultaneously. This usually involves some extended DNS functionality to be done right.
- Multiple links to the same provider, such as primary and backup links or two parallel load balanced links.

- Links to multiple providers, but using multiple static routes or a 3rd party solution (e.g. Radware LinkProof) to avoid full BGP routing.

4. Advanced BGP Techniques: Static Load Balancing

The previous section provided examples of basic activation of BGP Multihoming to two providers. These examples work fine for basic failover and redundancy, but they do not address some of the critical questions of how to load balance traffic across the two links. The general point is that BGP is just another routing protocol—it provides alternate paths, but does nothing by default to help you split your traffic sensibly between these paths. This section provides some "configuration snippets" and pointers for how to load balance traffic. We first take a look at outbound load balancing and then inbound load balancing, using standard BGP techniques.

4.1 Outbound Load Balancing

When you receive two different BGP routes (one from each provider) to the same destination prefix, there is a well-defined BGP Decision Algorithm for picking which route is best[8]. For standard eBGP routing, only ONE route can be active at any given time for a given prefix:

1. Do not consider iBGP path if not synchronized (this step usually disabled)
2. Do not consider route if there is no valid route to next hop
3. Pick the route with the highest weight (local to router)
4. Pick the route with the highest local preference (local to ASN)
5. Pick the route with shortest AS-path
6. Pick the route with lowest origin code (IGP < EGP < incomplete)
7. Pick the route with lowest MED
8. Prefer an eBGP route over an iBGP route
9. Prefer route with lowest next -hop metric
10. Tie breaker: route with lowest Router-ID is preferred.

Assuming no special modifications, the usual first decision point that decides which provider to use for a route is the length of the AS-path. AS-path is the number of "backbone hops", or administrative networks that must be traversed to get to the destination. Consider the following example:

```
router# sh ip bgp 65.242.124.0
BGP routing table entry for 65.242.124.0/24, version 74867653
Paths: (3 available, best #3, table Default-IP-Routing-Table)
  Not advertised to any peer

  [route #1]
  [as-path] 2914 10913 10913 10913 16919
    128.241.216.1 from 128.241.216.1 (129.250.59.3)
      Origin IGP, metric 10, localpref 100, valid, external
      Community: 2914:410 2914:2000 2914:3000

  [route #2]
  [as-path] 2914 10913 10913 10913 16919, (received-only)
    128.241.216.1 from 128.241.216.1 (129.250.59.3)
      Origin IGP, metric 0, localpref 100, valid, external
      Community: 2914:410 2914:2000 2914:3000

  [route #3]
  [as-path] 6128 701 10913 16919
```

[8] A good description and explanation of the BGP Decision Algorithm is provided at http://www.cisco.com/warp/public/459/25.shtml

```
    206.223.115.13 (metric 2530304) from 207.20.85.169 (207.20.85.169)
      Origin IGP, metric 10, localpref 100, valid, internal, best
```

Notice that 3 BGP routes are received, and only the last one with AS-path "6128 701 10913 16919" is selected "best", because the AS-path length is 4.

This kind of behavior may work OK if you get service from two roughly equal-sized providers. But more often then not, your outbound traffic will not be well-balanced. Consider the case of using one large Tier1 provider at an expensive price, and one Tier2 provider at a much cheaper bandwidth price. By default, the Tier1 network will generally have a shorter AS-path length to Internet destinations, since it is global and closer to the "core" of the Internet (through better peering). So without modification, your outbound traffic will prefer the more expensive provider.

Fortunately, there are tools to override default BGP behavior and define your own preferences to influence outbound traffic. The most common parameter used for this is BGP Local Preference, or "localpref." One common approach relies on the fact that you can move a lot of your traffic just by changes routing preferences for a few of the largest ASNs, rather than worrying about setting preferences for many small ASNs. The topology shown in Figure 4 illustrates the "Boulders and Pebbles" concept: *a small number of large networks usually account for the majority of your traffic.*

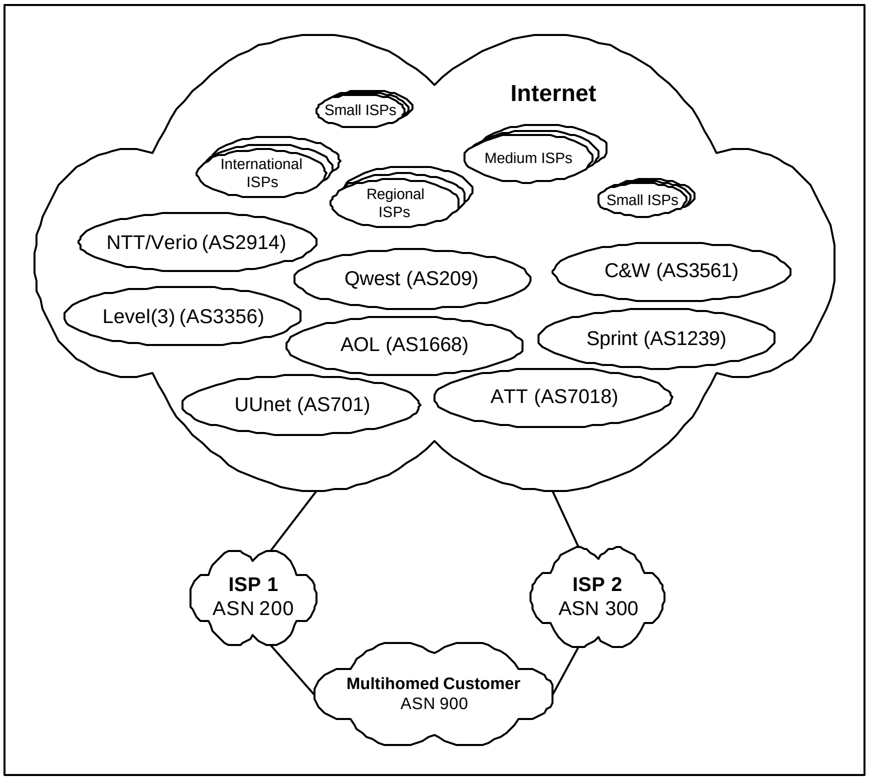


Figure 4. Major Internet Destination ASNs

Based on this observation, outbound load balancing can be achieved with the following steps:

1. Select a number of well-known large ASNs (excluding your directly attached providers). If possible, base this selection on actual traffic statistics. If not possible, expect to do some

experimentation with the steps below, and different ASN selections, until you achieve satisfactory results.

2. On the links you want to shift traffic to, use Cisco route-maps (or other vendors equivalent features) to match a subset of routes received from each provider, depending on whether one of these ASNs appears in the AS-path.
3. For these matching routes, increase the localpref to force these routes to be preferred.
4. Examine the resulting outbound traffic pattern, and add or delete more ASNs to shift outbound traffic until you are satisfied with the result.
5. Make sure to maintain this approach by periodically revisiting your choice of ASNs in step 1. Some ASNs may shrink over time, and others may grow (e.g. broadband providers in the U.S.)

While there are more scientific ways to do outbound load balancing, using techniques such as Cisco NetFlow analysis, these techniques all require some up-front measurement infrastructure to be developed (or purchased) and deployed[9]. In practice, the empirical approach presented here will probably be sufficient for many multihoming situations.

Here is a recipe for this approach, based on the network diagram used in Figure 3 in the previous section. Note that the ASNs chosen for this example are U.S.-centric. For international situations, you will need to consider the major ASNs for the region in question.

Router 1:

```
!
! Construct an AS-path access list to match major provider ASNs
! that appear anywhere in an AS-path (regular expression _ASN_)
! These are the provider ASNs that you wish to prefer through
! your link to ISP-1 (ASN 200).
!
! NOTE:  when applying this example to your real-life situation, make
! sure that none of the ASNs below match any of your directly attached
! providers, or this will not work!
!
! ASN 3561 is the main C&W ASN
! ASN 1239 is the main Sprint ASN
! ASN 7018 is the main AT&T ASN
! ASN 209 is the main Qwest ASN
! ASN 3356 is the main Level3 ASN
! ASN 7132 is the main SBC ASN
! ASN 6128 is the main Cablevision ASN
! ASN 22909 is a major Comcast ASN
ip as-path access-list 51 permit _3561_
ip as-path access-list 51 permit _1239_
ip as-path access-list 51 permit _7018_
ip as-path access-list 51 permit _209_
ip as-path access-list 51 permit _3356_
ip as-path access-list 51 permit _7132_
ip as-path access-list 51 permit _6128_
ip as-path access-list 51 permit _22909_
!
! Now create route maps, which apply these AS-path access lists.
! Remember, default localpref is 100, and higher means more
! preferred.
!
```

[9] One popular off-the-shelf tool for integrated BGP and traffic analysis is from Adlex, http://www.adlex.com, although many route control products such as netVMG offer similar capabilities. If you do not mind using UNIX-based open source packages, take a look at a combination of cflowd (http://www.caida.org/tools/measurement/cflowd: an open-source NetFlow data collector) and flowscan (http://net.doit.wisc.edu/~plonka/FlowScan: a display and graphing tool for cflowd data).

```
! Set higher localpref for subsets of ASNs through ISP-1
route-map pref-ISP1-outbound permit 10
 match as-path 51
 set local-pref 200
! Make sure to match remaining routes with default localpref 100, so
! they are still valid through ISP-1, even if not preferred.
route-map pref-ISP1-outbound permit 20
 set local-pref 100
!

! Now don't forget to apply these route maps to each peering session
! and do a "clear ip bgp <x.x.x.x> soft in" to reset.
! Remember, these are activated "in" since they apply to routes you
! are receiving from your providers.
! Since local-pref is a transitive attribute, the preferences set by
! these route maps will propagate to all your iBGP routers, if you have
! more than one.
!
router bgp 900
 neighbor 1.1.1.1 route-map pref-ISP1-outbound in
!
end

router# clear ip bgp 1.1.1.1 soft in
```

Router 2:

```
!
! Construct AS-path access list to match intermediate ASNs you
! wish to prefer through your link to ISP-2 (ASN 300).
!
! ASN 1 is the main former Genuity/BBN ASN
! ASN 1668 is a large ASN used by AOL
! ASN 2914 is the main ASN used by Verio/NTT
! ASN 701 is the main backbone ASN used by UUnet
ip as-path access-list 52 permit _1_
ip as-path access-list 52 permit _1668_
ip as-path access-list 52 permit _2914_
ip as-path access-list 52 permit _701_
!
! Set higher localpref for subsets of ASNs through ISP-2
route-map pref-ISP2-outbound permit 10
 match as-path 52
 set local-pref 200
! Again, accept remaining routes, just with default localpref
route-map pref-ISP2-outbound permit 20
 set local-pref 100
!
router bgp 900
 neighbor 2.2.2.1 route-map pref-ISP2-outbound in
!
end

router# clear ip bgp 2.2.2.1 soft in
```

There are a few variations to this approach. One common variation is to:

1. Start by preferring all traffic through ISP-1 by setting all ISP-1 routes to local-pref 150. You should now be sending ALL outbound traffic through ISP-1, and nothing through ISP-2.

2. Move chunks of traffic over to ISP-2 one at a time, using the AS-path matching technique in the above example.
3. Iterate over enough candidate ASNs until you achieve the desired load-balancing.

4.2 Inbound Load Balancing

Unlike outbound load balancing, where you are in full control of which link you prefer to send traffic out on, inbound load balancing is an even less-exact science. We will discuss a few of the most common methods of controlling this, based on how you tune your routing announcements out to each provider. However, none of these techniques is foolproof. While these techniques work more often than not, in the end, it is up to end users' networks to decide how to route traffic back to you. They may choose to override what they learn from your routing announcements if they have a conflicting routing policy.

The types of inbound load balancing we will discuss are:

- Prefix Engineering: The practice of breaking an aggregate announcement up into more specific announcements, and selectively announcing the more specifics.
- AS-prepending: The practice of lengthening your AS-path through selected providers by prepending your own ASN one or more times to the AS-path of your announced route.
- Provider-supported BGP Communities: The practice of using BGP community tags, if your providers support it, to influence how they propagate your routing announcements.

Prefix Engineering: This technique works when you have a large enough address block to break it into pieces, each bigger than a /24. A typical example is a customer with a /19. They can announce the /19 route to ISP-1 and ISP-2, upstream providers. Then, they can announce a more specific /20 route through ISP-1. Since only ISP-1 propogates this more specific route, and since more specific routes are always preferred over larger aggregates, ISP-1 will tend to carry all inbound traffic for the /20 block. The customer can repeat this with the other /20 in their /19 block, through ISP-2, and try to split inbound traffic between the two ISPs. This assumes that they are numbering servers in their site equally between both /20's.

Prefix engineering is one of the major features used when customers have multiple sites, and wish to draw inbound to a particular site. You need to plan ahead for this, to assign address blocks from your aggregate to each site according to your specific route announcements.

Here is an example of prefix engineering for inbound traffic control, based on Figure 3 in the previous section (this assume PI-space):

- Customer ASN 900 will announce /19 to ISP-1, as well as the lower half as a more specific /20.
- Customer ASN 900 will announce /19 to ISP-2, as well as the upper half as a more specific /20.
- Customer can watch the impact this has on inbound traffic shift, and iterate by tuning the size of the more specific announcement.

AS-Path Prepend: Since traffic will route inbound to you based on the BGP routes they receive, one attribute you can control is the length of the AS-path. You can choose to prepend your ASN multiple times before you announce your prefixes out to a given provider. You can also combine Prefix Engineering and AS-Prepend. Here is one scenario, based on Figure 1 in the previous section (this assumes PA-space), plus the assumption that the customer has two separate /24 PA-space allocations:

- Customer ASN 900 will announce first /24 through ISP-1 normally, but prepend it's own AS 3 times when announcing the second /24 through ISP-1.
- Customer ASN 900 will announce second /24 through ISP-2 normally, but prepend it's own AS 3 times when announcing the first /24 through ISP-2.

- This has the impact of preferring ISP-1 as the inbound path for the first /24, and preferring ISP-2 as the inbound path for the second /24.
- Customer can watch the impact this has on inbound traffic shift, and apply limited tuning by changing the number of AS-path prepends.

Provider-Supported BGP Customer Communities: Some providers have developed rich sets of communities they will accept from customers, which influences the behavior of how they treat the route they receive from you, both internally in their own network, and externally as they propagate this route to their peers. To take advantage of this, you must:

1. Investigate the set of customer-usable communities that your provider supports.
2. Make sure your BGP session is enabled to send communities to your provider, and that they are enabled on their end to receive these communities.
3. Set up route maps that match the right prefixes, and set the right communities to these prefixes.
4. Apply these route maps outbound on the appropriate provider BGP sessions, and reset these sessions (“`clear ip bgp <x.x.x.x> soft out`” if your provider accepts soft reconfiguration from you, otherwise you must resort to a hard reset “`clear ip bgp <x.x.x.x>`”).
5. Use your provider’s Looking Glass or route views servers to ensure that the community settings are being received and are having the desired effect.

Note that there is a wide variance in provider support for BGP Customer Communities. Some providers do not support any communities at all, many providers support basic communities, and a few providers support some very advanced BGP Customer Community features. The types of features that may be available via provider-supported BGP communities are:

- *AS-path Prepending:* Your provider will respond to routes tagged with this community by prepending its own AS a variable number of times. Usually communities are available to set the prepend to a value between 1 and 5 times, so that you have room to tune this parameter to your taste.
- *Selective AS-path Prepending:* Very powerful feature for inbound traffic control. Your provider may allow you to control how they prepend to specific upstream ASNs (usually just major providers). For example, you may be able to set communities to tell your provider to prepend 2 times when announcing to UUnet, 3 times when announcing to Sprint, but only once when announcing to Level3.
- *Local-Pref:* You can tag routes with this family of community values to influence the local-pref values your provider attaches to your route announcements within its network. For instance, for cost reasons you may want to force ISP-2 to de-preference its direct inbound path to you, and force it to route your inbound traffic through its peering with ISP-1 to get to you.

More Examples:

- http://www.nanog.org/mtg-0110/smith.html has some excellent examples of AS-path Prepend, Prefix Engineering, and Provider-support Communities techniques and configurations.
- If you want to see more examples included in this white paper, please contact the authors. We are happy to respond to feedback and produce a new revision.

5. Multihoming Strategies and Technologies

The previous section included several advanced techniques for outbound and inbound load-balancing using standard BGP features. However, there are a number of limitations to standard BGP. Two of the biggest are that:

- BGP is not sensitive to cost of bandwidth
- BGP is not sensitive to performance.

Decisions are based on static BGP selection parameters that you configure into your router. This can make it difficult to achieve important performance and cost goals. This section describes new route control technology that was designed to enhance BGP Multihoming, and make a series of new performance and cost optimizations possible.

Other drawbacks that are commonly considered when evaluating BGP Multihoming are the cost of local loops, the hassle of negotiating competitive ISP bandwidth contracts (and the risk of getting stuck in a bad contract if you don't), the inability to easily change providers once they are provisioned, and the equipment costs and engineering resources required to maintain multiple ISP connections.

This section discusses these pitfalls, and for each one, presents strategies to overcome them and design the most optimal BGP Multihoming solution, from both a cost and performance standpoint.

5.1 Route Control: Optimizing Cost and Performance

Route Control has emerged as a way to address the weaknesses of standard BGP Multihoming. Route Control is available both as a service and as a network appliance[10]. To illustrate route control, consider the following hypothetical requirements that cannot be addressed by standard BGP. These are based on the examples from Section 3, an example customer who is multihoming to ISP-1 and ISP-2:

- ***Identify and Avoid Providers' Backbone or Peering Weaknesses:*** A portion of the routes that the customer statically prefers through ISP-1 have latency that is three to four times worse than going through ISP-2. Conversely, some of the routes the customer statically prefers through ISP-2 could be improved significantly through ISP-1. Customer wants to prefer the best performing ISP for each Internet destination, but does not have the resources to continually monitor per-destination performance, or maintain and dynamically update resulting per-prefix BGP preferences.
- ***Take Advantage of Cheaper Providers Without Risking Poorer Performance:*** ISP-1 is significantly cheaper than ISP-2, which is a large, premium provider. But ISP-1 has a poorer reputation for quality and uptime. The customer wants to use ISP-1 to save money. But if it is not providing good performance to some routes, the customer wants to dynamically re-route any impaired routes through ISP-2. As soon as performance improves again on ISP-1, these routes should revert back, to avoid triggering high 95th percentile utilization bills through ISP-2.
- ***Tune My Traffic Load Balancing Based on Provider Costs and Commits:*** The customer has minimum bandwidth commits to ISP-1 and ISP-2. The customer wants to send enough traffic out each ISP to meet (but not exceed!) these commits. Any extra traffic, should be sent over the cheaper ISP. This approach will maximize savings on bandwidth costs.

Route control technology has matured throughout 2002, and has been successfully deployed by customers to lower their bandwidth costs, increase their Keynote™ ratings, and get real-time visibility into traffic bottlenecks. Some of the key features you should expect from adopting this technology are:

1. *Per-route, Per-Provider Performance Monitoring:* Automatically measures the performance through each of your providers, in real time, for each route that is active to your site. Provides your router with optimized routing information using standard BGP updates.
2. *Tracks and Reacts To Provider Bandwidth Bills:* Keeps continual track of bandwidth usage through each provider and compares it against your provider's billing methodology to implement complex cost-control strategies. For example, it can fulfill minimum commits through each provider, then route any extra traffic through the cheapest provider.

[10] For example, Equinix offers an Intelligent Routing Service based on products from vendor netVMG (www.netvmg.com). Sockeye and Route Science are other examples.

3. *Detailed Reporting:* Provides reporting of all measured values through a web portal, to report on bandwidth utilization, provider performance, and route change activity that is initiated in response to network problems, and the degree of performance improvement gained by such activity.
4. *Policies That Support Cost and Performance Improvement:* Perfoms a complex trade-off of cost versus performance: keeps traffic flowing along cheapest providers if performance is acceptable, but can reroute traffic to other providers in response to major performance impacts, such as provider failures, congestion caused by incidents such as Internet worm outbreaks, and maintenance windows. Can also be set for cost-only, or performance-only, optimization modes.

5.2 ISP Contracts and Bandwidth Billing

Bandwidth contracts have been evolving quickly in most major ISP markets. You need to keep abreast of contract expectations, and learn what kinds of contracts are helpful or harmful for implementing BGP Multihoming. In addition, it is important to understand 95th percentile bandwidth billing. While this type of billing provides the flexibility to pay only for bandwidth consumed, there are traffic situations you need to avoid that can needlessly multiply your bandwidth costs.

The Art of ISP Contract Negotiation: ISP contracts impact both your cost structure and your flexibility. When you multihome, you have the ability to control which ISP carries the bulk of your traffic (and you may want to change this preference based on your experience), so it is important to avoid signing up for high traffic commitments to any single provider. In general, you should seek low traffic commits through services that provide burstable 95th percentile billing, for which you only pay for traffic consumed. When you purchase services using LAN-style interfaces (Fast Ethernet and Gigabit Ethernet), it is not uncommon for the minimum commit to be at least 10% of the port speed.

You are not dependent on a single ISP when multihoming, so you will want to avoid long contract terms in case you decide to replace one of your providers. If you are located in a carrier-neutral data center, you will also want to avoid long contract terms to take best advantage of price competition among participating ISPs. Some providers allow month-to-month terms at nearly the same prices as their 12-month or longer terms.

You should also be aware of unfriendly billing tiers. Some providers may charge you a higher rate if you burst above your bandwidth commit for even one month, and lock you into that higher tier for a number of months before you are allowed to decrease to a lower usage tier. If possible, avoid this type of tiered billing.

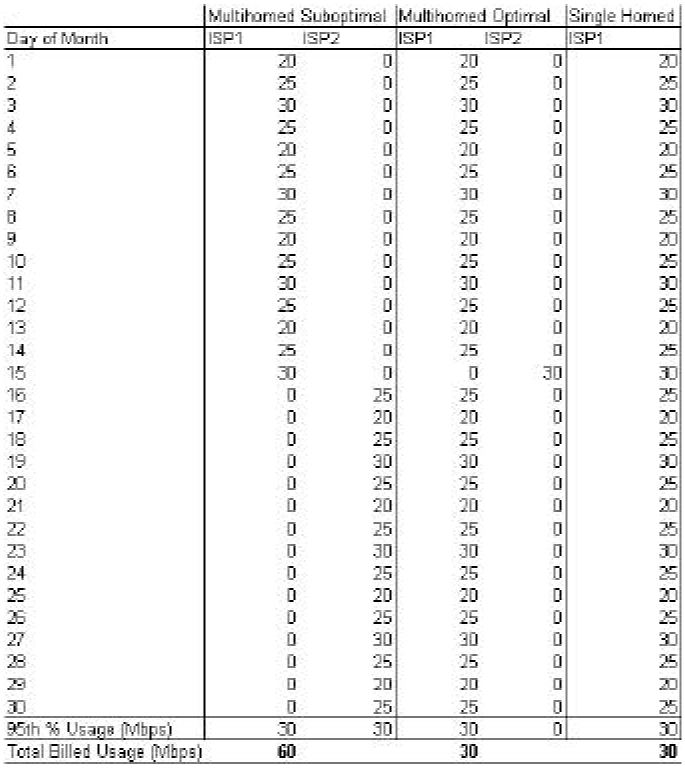
Some carrier neutral data centers re-sell the transit services of the ISPs that are collocated at competitive rates. This may be a useful way to explore pricing available from a number of providers through a single request. You can also use this arrangement to simplify your contract negotiations, and consolidate your bills for all your services.

Selecting ISPs That Fit You: If you have particular bandwidth profiles, look for ISPs that are architected to better meet your needs. For instance, some ISPs have high quality capacity and routes to Asia, but a very thin U.S. footprint. Other providers have a strong U.S. backbone, or lots of capacity and comprehensive routes to European countries. In the U.S., some providers specialize more in connecting business, while other networks are heavily focused on providing broadband DSL or Cable access to residential consumers. Some providers focus on reselling inexpensive, commodity Internet access, while others focus on combining access to multiple Tier1 providers to provide a high-value service. If you have special bandwidth needs, don’t make the mistake of assuming all Internet access is equal. Request the network architecture, routing and capacity information from ISPs so you can make an informed choice.

95th Percentile Billing--Avoid Double Bandwidth Bills: One of the pitfalls to avoid when multihoming is doubling your bandwidth bills by sending all your traffic through one provider for part of the month, then switching your traffic to prefer another provider for the remainder of the month. Each ISP will bill you for the entire 95th percentile load for the month. To understand why, let’s look at how 95th Percentile Billing works.

95th Percentile Billing is most common when you buy burstible access to providers. Usually, a provider takes a sample of your bandwidth utilization every five minutes, for the entire month. At the end of the month, they sort these samples, throw away the highest 5%, and bill you for the remaining peak. For an average 30-day month, this gives you 36 hours of “free” bandwidth peaks that are thrown away. This can be a huge advantage when you only need to fail over to an expensive provider to avoid a 4-hour catastrophe on your cheaper ISP link—you never pay the expensive provider for the 4 hours of peak traffic you sent through it! But if you send traffic through a provider for longer than 36 hours (greater than 5% of the month), you will trigger a bill for that bandwidth for the entire month! It is critical to be aware of this before shifting large amounts of your traffic between providers mid-month[11].

Figure 5 is a tabular example of bandwidth usage, which shows optimal and suboptimal use of 95th Percentile billing principles (simplified to show daily bandwidth samples, not 5-minute samples). In the Suboptimal case, a customer changes preference from one provider to another mid-month, *which doubles the billed bandwidth*. In the Optimal case, ISP2 is only used for one day (e.g. due to an outage or degradation in ISP1). This usage is for less than 5% of the month, so it is discarded and the total billed bandwidth is the same as the single-homed case.



	Multihomed Suboptimal		Multihomed Optimal		Single Homed
Day of Month	ISP1	ISP2	ISP1	ISP2	ISP1
1	20	0	20	0	20
2	25	0	25	0	25
3	30	0	30	0	30
4	25	0	25	0	25
5	20	0	20	0	20
6	25	0	25	0	25
7	30	0	30	0	30
8	25	0	25	0	25
9	20	0	20	0	20
10	25	0	25	0	25
11	30	0	30	0	30
12	25	0	25	0	25
13	20	0	20	0	20
14	25	0	25	0	25
15	30	0	0	30	30
16	0	25	25	0	25
17	0	20	20	0	20
18	0	25	25	0	25
19	0	30	30	0	30
20	0	25	25	0	25
21	0	20	20	0	20
22	0	25	25	0	25
23	0	30	30	0	30
24	0	25	25	0	25
25	0	20	20	0	20
26	0	25	25	0	25
27	0	30	30	0	30
28	0	25	25	0	25
29	0	20	20	0	20
30	0	25	25	0	25
95th % Usage (Mbps)	30	30	30	0	30
Total Billed Usage (Mbps)	**60**		**30**		**30**

Figure 5. Optimal and Suboptimal Impact of Multihoming On 95th Percentile Billing

Whether you use route control technology, or simply make wise choices on when and how to change your BGP load balancing policy, you should be able to implement multihoming without significant increases to your billed bandwidth.

[11] An excellent graphical example of 95th percentile “bill multiplication” due to inefficient traffic patterns is found in the “Economics of Multihoming and Route Control” white paper available at http://www.netvmg.com/cgi-bin/registration.cgi?id=2

5.3 Advantages of Neutral Colocation

There are some natural advantages to multihoming within a neutral colocation model:

- ***Eliminate Local Loop Costs:*** One of the biggest advantages of neutral colocation is eliminating the cost of local loops to reach your Internet providers. For content providers, web hosters, and bandwidth resellers, this can make or break their bandwidth margins. Even for enterprises that need to backhaul their Internet traffic back to a nearby headquarters, there can be cost advantages: (1) through consolidating multiple Internet local loops into a larger backhaul circuit, (2) through access to more competitive local loop prices for the backhaul circuit, (3) through deploying front-end caching and other bandwidth reduction strategies within the neutral colo, prior to backhauling traffic.
- ***Choose from a Longer List of Providers:*** Due to their role as inter-provider traffic exchange points, neutral colocation providers usually have a long list of ISPs present. You will usually be able to choose from multiple providers that match your requirements.
- ***Easily Switch Providers:*** Since cross-connects to ISPs at neutral facilities can be changed in a matter of hours or days, you have the flexibility to easily add, change, or drop providers, to react to changing requirements, take advantage of better contract terms, or avoid a poor performer.
- ***Benefit from Competitive Pricing and Contracts:*** Neutral colocation providers create a competitive marketplace for purchasing bandwidth. There is a lower cost and provisioning complexity to acquire customers in this environment. ISPs that have core inter-provider peering at the neutral facility may have lower costs to carry your traffic, since they can often directly hand it off to a destination network before the traffic leaves the facility. Whatever the reasons, you will probably be able to get more flexible bandwidth commits, shorter term contracts, and better prices at provider-neutral data centers than you would when procuring bandwidth to your office location.
- ***Robust Power, Cooling, Security, and Fiber Redundancy:*** Neutral colocation centers may offer facilities robustness and scalability that is expensive to duplicate on a smaller scale in corporate data centers.

5.4 Multihoming Platforms

Even within a neutral colocation center, implementing BGP Multihoming requires you to deal with multiple providers, and dedicate a router port to each one. To keep the process manageable, most companies that multihome only choose two or three providers. Even when the cost of switching providers is low, there is technical reconfiguration that requires change management and operational planning.

Some neutral colocation providers have recently developed “transit exchanges”—multihoming platforms that allow enterprises and content providers to access multiple ISPs through a single router port. Common features of this type of platform are month-to-month terms, no bandwidth commits, competitive pricing, and web-based reconfiguration of providers each month[12].

[12] One example of a recently announced multihoming platform is Equinix Direct, which is described at http://www.equinix.com/prod_serv/network/ed.htm

6. Conclusions

Multihoming, when implemented well, can help companies meet their Internet performance, reliability, and redundancy goals, without increasing their costs. It also helps companies reduce their business reliance on a single provider, giving them dramatically greater opportunities for bandwidth cost control and contract flexibility.

The purpose of this white paper is to help companies evaluate BGP Multihoming as an Internet access choice, review some of the basic concepts and router configurations used to implement it, and discuss more advanced techniques and strategies for getting the most out of BGP Multihoming.

This is a living document, and the authors will continue to update it periodically to reflect new developments.

A. References

http://www.nanog.org/mtg-0110/smith.html - A comprehensive collection of slides developed by Cisco to address BGP Multihoming.
http://avi.freedman.net/fromnetaxs/bgp/bgp.html - A somewhat dated (1997), but useful, writeup on BGP Multihoming, from a seasoned ISP architect who has lived through it.
http://www.ipsyn.net/cisco/tutorials/workshops/isp-workshop/WhitePapers/Multihoming.PDF - A somewhat dated white paper on BGP Multihoming presented by Cisco.
http://www.cymru.com/Documents/secure-bgp-template.html - A very useful and frequently updated template for implementing robust and secure BGP sessions to other providers.
http://www.netvmg.com/products/resource_library.html - A collection of white papers regarding BGP Multihoming planning and optimization from a route control vendor.
http://joe.lindsay.net/bgp.html - A useful set of links for learning about BGP in general, but does not appear to have been updated since Sept 2001.
http://www.cisco.com/warp/public/459/bgpfaq_5816.pdf - Cisco's BGP Support FAQ, practical answers to frequent BGP support questions.
http://www.cisco.com/en/US/tech/tk826/tk365/tk80/tech_protocol_home.html - Cisco's BGP Routing Protocol main page.
http://isp-lists.isp-planet.com/isp-bgp - Web archives of the ISP-BGP mailing list.
http://www.networkingunlimited.com/whitepapers/white008.pdf - Multihoming to Two ISPs white paper by Dr. Vincent Jones, author of "High Availability Networking with Cisco", Addison-Wesley.