Multi-Homing

May 2004

Absolute Availability

The Economist in its article on the online economy in May 2004 reported that the 200 million Americans who now have online web access are likely to spend more than $120 billion in 2004 through online services. There are large investments being made in creating online service platforms, and the builders of such platforms are looking for one common aspect from their technology provider - 100% high performance availability. Not the so called "5 9's" availability of 99.999% availability, but an absolute commitment of availability all of the time.

Simple to state as a requirement, or course, but perhaps not quite so easy to construct.

This is an important requirement from one of the more critical customer segments of the Internet, so perhaps its worth spending some time in looking at how this requirement is approached today and what our options are in the future.

Why is multi-homing a topic of deep interest in the Internet industry?

It seems almost contradictory that we should be considering this topic at all. If we head back to the early 1960's and the research into packet switched networks, Paul Baran's work looked at concepts of network design where the network was more resilient than individual components. The concept centered around the ability of the network to detect and 'heal' component failure, and to do so without disrupting active connections that were transiting the network. The IP architecture has largely achieved these objectives, and with a rich enough underlying topology it is possible in IP to have seamless network healing in the face of component failure.

So if IP really is as resilient as is claimed then what's the problem here? Well the story is not quite so bright as that, and there always appear to be critical single points of failure. There may be a single connection from the server platform to the network service provider. Well, install two circuits then. The circuits may terminate on the same access equipment, or may use a common component in the same access site. Well, install each access circuit to distinct equipment, possibly in different sites. So we now have a service provider that uses a highly resilient architecture that has dual paths from access nodes through each POP, with two or more POPs in each major location, a core infrastructure that has multiple paths across the network, with multiple external transit relationships, also provisioned in a highly resilient fashion. Are we there yet for 100% availability? Unfortunately not. There is still the issue that a single network domain has a single routing state, and it has been known for the routing system itself to get wedged into states that isolate customers. So if you want absolute availability then maybe you also need to use multiple service providers, and "multi-home" your service platform to both providers at once.

Its not the only reason to multi-home of course, and there is also the issues of load balancing, tariff optimisation and performance optimization that are options when your local services are connected to the network through two or more providers.

Multi-homing Today

So given that we want to construct an environment where a local network is connected through two or more upstream network service providers, then what's the best way of doing this?

The 'classic ' architecture of multi-homing today is that the multi- homed site needs to advertise its address block to all its transit service providers, who in turn advertise the addresses onward to their interconnected peers, and so on. So what you need to construct a multi-homed site is three things:

- your own address block (or a block of 'provider independent' space),
- your own Autonomous System number, and
- multiple upstream network connections.

The Autonomous System (AS) number is optional, although many network administrators tend to regard multi-origin address blocks with some level of suspicion as a configuration error, so perhaps obtaining an AS number is prudent in this situation.

Its even optional to have your own block of provider independent address space, although it is highly advisable that you should use a unique block of independent space. While it is possible to use addresses provided from one upstream's address aggregate and advertise the fragment to the other providers, you may find that this is a less than optimal solution. Wherever the fragment is propagated in the Internet the incoming traffic will follow the path of the more specific fragment, rather than the path defined by the aggregate announcement. How then do you get incoming traffic to take the path of the original address provider if that is your preferred policy? Also many providers tend to frown on attempts to fragment their aggregate address, as this is not exactly a friendly act in terms of the Internet's routing space, so it may also be the case that other providers may not accept your advertisement of an address fragment from another provider. Also, if you cancel your service contract with the provider from whom you are drawing your address block then renumbering is a forced consequence. And, as the saying goes, 'renumbering is **hard**!'

So, armed with your address block, your AS number and multiple upstream connections, what's next? Each upstream connection is supported by an eBGP session, you announce your address block to each provider, and receive their routing table. You are now multi-homed. No protocol changes, no application changes, nothing - apart from just one consideration - you have just added another entry into the Internet's inter-domain routing system.

Multi-homing via Routing

So this approach, as used in IPv4 for multi-homing support, preserves the semantics of the IP address as both an endpoint identifier and a forwarding locator. For this to work in a multi-homing context it is necessary for the transit ISPs to announce the local site's address prefix as a distinct routing entry in the inter-domain routing system.

The local site's address prefix may be a more specific address prefix drawn from the address space advertised by one of the transit providers, or from some third party provider not current directly connected to the local site. Alternatively, and preferably, the address space may be a distinct address block obtained by direct assignment from a Regional Internet Registry as Provider Independent space. Each host within the local site is uniquely addressed from the site's address prefix.

All transit providers for the site accept a prefix advertisement from the multi-homed site, and advertise this prefix globally in the inter-domain routing table. When connectivity between the local site and an individual transit provider is lost, normal operation of the routing protocol will ensure that the routing advertisement corresponding to this particular path will be withdrawn from the routing system, and those remote domain domains who had selected this path as the best available will select another candidate path as the best path. Upon restoration of the path, the path is re-advertised in the inter-domain routing system. Remote domains will undertake a further selection of the best path

based on this re-advertised reachability information. Neither the local or the remote host need to have multiple addresses, nor undertake any form of address selection.

The path chosen for forward and reverse direction traffic flows is a decision made by the routing system. However there are an increasing number of configuration options that allow the site to not only achieve full failover to alternate paths when there is connectivity failure, but also, when there are alternate paths available, to perform various forms of traffic engineering to optimise performance, cost or other policy-related objectives. Outgoing traffic may be biased by local preference settings applied to learned routes. Incoming traffic paths may be altered if neighbouring domains support communities that allow preference setting via community values. And of course there's always AS path prepending, and, as a last and definitely ill-advised resort, there's always the selective advertisement of more specific address prefixes along specific preferred paths. In this context multi-homing has matured sufficiently that it's possible to engineer a highly resilient service solution that not only achieves high availability requirements but also allows for various forms of load balancing, cost optimization and policy constraints. An impressive outcome, particularly considering that even this additional functionality can be achieved without changes to the IP protocol, the DNS or applications.   This approach could be used in an IPv6 context, and, as with IPv4, no modifications to the IPv6 architecture are required to support this approach.

The Limits of Multi-homing

So how many sites can be multi-homed?

Each site that multi-homes in this fashion adds a further entry in the global inter-domain routing table. So, in using this approach, the number of multi-homed sites is limited by the number of entries that we can add into the Internet's routing system.

Within the constraints of current routing and forwarding technologies it is not clearly evident that this approach can scale to encompass a population of multi-homed sites of the order of tens or hundreds of millions of such sites. The implication here is that this would add a similar number of unique prefixes into the inter-domain routing domain, which in turn would add to the storage and computational load imposed on routing elements within the network. This scale of additional load is not supportable within the current capabilities of the switching elements of the global Internet, nor is it clear at present that the routing capabilities of the entire network could be expanded to manage this load in a cost-effective fashion, within the bounds of the current inter-domain routing protocol architecture.

It appears that the current approach generally meets functional criteria for multi-homing approaches with one noteable exception: scaleability. And if the Internet can be summarized in a single word, that would have to be "scale".

Alternative Approaches to Multi-Homing?

So it would appear that we are looking towards IPv6 with a view to supporting a truly massive deployment of end systems, where the unit of service is in the thousands of millions. Yet we have to accept that do not have an approach to scaling routing to a similar level that does not make extensive use of hierarchy and topology to reduce the routing information load.

Also, if we see multi-homing as a common approach to achieving cost- effective service resilience, then we should expect to see multi- homing encompass much more than thousands, and probably more than millions of sites over time.

Combining these observations it seems we have some work to do with multi-homing. We'd like to see a scaleable approach to multi-homing. Somehow we need to provide the functional outcome of service resilience that routing-based multi-homing already offers, but without the routing overhead that is associated with the current approach.

This makes multi-homing very challenging indeed. In the next column I'll take a look at the general approaches that might be capable of meeting this objective in the context of IPv6 in particular. The interesting observation is that the alternatives in this space are no longer small-scale changes to the protocol architecture - they cut to the very heart of the architecture and appear to require significant changes in the underlying assumptions behind the IP design.

I trust I've raised your interest enough to read on next month to see what such alternatives may be ….

Geoff Huston

Disclaimer

About the Author

GEOFF HUSTON holds a B.Sc. and a M.Sc. from the Australian National University. He has been closely involved with the development of the Internet for the past decade, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. Huston is currently the Chief Scientist in the Internet area for Telstra. He is also the Executive Director of the Internet Architecture Board, and is a member of the APNIC Executive Committee. He was an inaugural Trustee of the Internet Society, and served as Secretary of the Board

of Trustees from 1993 until 2001, with a term of service as chair of the Board of Trustees in 1999 and 2000. He is author of a number of Internet-related books.