Six/One Router: A Scalable and Backwards Compatible Solution for Provider-Independent Addressing

Christian Vogt
Ericsson
Hirsalantie 11
02420 Jorvas, Finland
+358 9 299 2142
christian.vogt@ericsson.com

ABSTRACT

The scalability of the Internet routing system suffers from an increasing demand for provider-independent, non-aggregatable IP addresses in networks at the Internet edge. New routing architectures have been proposed that mitigate this problem through indirection between provider-independent addresses at the edge and aggregatable, provider-allocated addresses in the core of the Internet. A major challenge in these architectures is backwards compatibility. Address indirection requires support at the sender and the receiver, so without appropriate backwards compatibility support, it defeats communications between the upgraded and the legacy Internet. This paper proposes an address-indirection-based solution that is backwards compatible. The solution is shown to offer the benefits of provider-independent addressing in a scalable and backwards compatible manner, and to provide the incentives necessary to foster its early deployment.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design – *network communications, network topology.*

General Terms: Design.

Keywords: Routing, Multi-homing, Scalability.

1. INTRODUCTION

Networks at the edge of the Internet increasingly often switch from classic, provider-allocated IP addresses to provider-independent IP addresses. They do this to avoid internal re-addressing when changing providers and, when multi-homed, to facilitate load balancing and fail-over between providers. Unfortunately, this development causes undesirably fast growth in the size and update frequency of the global routing table, because provider independence defeats the aggregatability of IP addresses, and multi-homing-related traffic redirection causes routing table updates Internet-wide. Both strains memory and processing capacities in Internet core routers to an unacceptable extent [RAWS].

To enable the use of provider-independent IP addresses in edge networks without adverse effects on global routing scalability, *address indirection* [Fa2007, Jen2007] has recently received much attention. Provider-independent *edge addresses*, for use within edge networks, are hereby mapped to provider-allocated *transit addresses*, for use in the Internet core. Transit addresses are listed in the global routing table and are hence globally routable, whereas edge addresses are routable only locally within their edge network. New *indirection routers* transform packets from edge to transit addresses at the border of the originating edge network, and back to edge addresses at the border of the receiving edge network. This is transparent to hosts in both edge networks. A supplementary system for *mapping resolution* supplies indirection routers with the mappings they may need for remote edge addresses.

A challenge with address indirection is backwards compatibility: Since edge addresses are not globally routable, packets sent to one of them are dropped at the border of the originating edge network if not transformed by an indirection router. Edge networks that adopt address indirection thus become unreachable from legacy networks. The problem is exacerbated when address indirection is realized through tunneling because the extra IP header makes packets sent to legacy edge networks unprocessable by the receiving host. New infrastructure has been proposed [Lw2007] to proxy indirection router functionality for legacy edge networks. But this lacks convincing deployment incentives, because the deployment and operational costs for the new infrastructure must be born by providers that obtain little benefit from it.

But this is costly, and its required installation close to legacy edge networks limits its chances for sufficiently wide deployment.

This yields four design goals for new routing architectures:

- *Scalability* – Addresses that are used for routing across the Internet core must be aggregatable. Provider load balancing and fail-over of multi-homed edge networks must be transparent to the Internet core.
- *Re-addressing avoidance* – The addresses used in edge networks must be provider-independent so that they can be retained after provider changes.
- *Multi-homing support* – Multi-homed edge networks that use provider-independent addresses today can redirect traffic between their providers. These redirection capabilities must be retained.
- *Incremental deployability* – Success of Internet enhancements requires backward compatibility and incentives for early adoption. Both must be provided.

This paper presents Six/One Router, an address indirection protocol that satisfies these requirements based on two key concepts:

MobiArch'08, August 22, 2008, Seattle, Washington, USA.

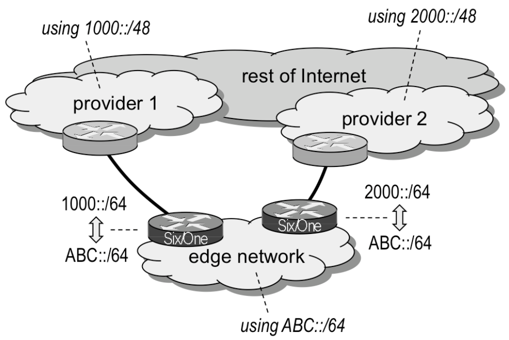

Figure 1: Multi-homed edge network with Six/One Router support

- *One-to-one address translation* — Indirection routers translate packets from edge to transit addresses such that the packets do not necessarily have to be translated back to edge addresses. Each edge address maps onto one unique transit address per edge network provider, and each transit address maps onto a unique edge address. To assist a remote indirection router at the receiving edge network, in case one exists, in translating transit addresses back to edge addresses, indirection routers add auxiliary information to outgoing packets using a header extension. The remote indirection router removes the header extension before forwarding the packets to the recipient host.
- *Reachability at transit addresses* — Hosts in edge networks with Six/One Router support can be reached at transit addresses in addition to their edge addresses. This makes them reachable from legacy edge networks. The hosts themselves still use their edge addresses. A local indirection router translates between local edge and transit addresses, while leaving the addresses from the remote legacy edge network untouched.

As a result, Six/One Router leaves packets processable by hosts if performed on only one side of a packet exchange. When performed bilaterally, Six/One Router turns into an equivalent to tunneling and retains the end-to-end semantics of edge addresses. Since edge addresses are globally unique and unambiguously identifiable given a transit address, Six/One Router avoids issues that are commonly associated with the traditional use of address translation [Sr2001] in circumventing address shortage and address allocation hurdles in IPv4.

Like other address indirection protocols, Six/One Router depends on a system for mapping resolution. Six/One Router is compatible with existing mapping resolutions systems, such as DNS Map [Vo2008b], NERD [Le2008], APT Default Mappers [Jen2007], CONS [Br2007], and ALT [Fa2007a].

In the remainder of this paper, the protocol operation of Six/One Router is described in section 2 and evaluated relative to the aforementioned design goals in section 3.

2. PROTOCOL OVERVIEW

This section motivates and describes the key components of Six/One Router.

2.1 Network Setup and Address Allocation

An edge network that deploys Six/One Router – it will be called *upgraded* edge network henceforth as opposed to a *legacy* edge network – owns a set of provider-independent edge addresses, and it is allocated a set of transit address by each of its providers. Edge addresses are for use within this and other upgraded edge networks. They are globally unique, but do not appear in the global routing table because they cannot be efficiently aggregated. For global reachability, every edge address is instead mapped to a transit address from each provider. Indirection routers with Six/One Router support, so called *Six/One routers*, translate between edge and transit addresses. Upgraded edge networks deploy a Six/One router on every *border link* that connects them to a provider.

Figure 1 illustrates this setup and addressing for an edge network that is multi-homed with two providers. This edge network has one border link to each provider, so two Six/One routers are in use. The edge addresses are from the prefix **`ABC::/64`**. The Six/One router on the border link to provider 1 translates between those and transit addresses from the prefix **`1000::/64`**; the Six/One router on the border link to provider 2 translates between the same edge addresses and transit addresses from the prefix **`2000::/64`**.

Throughout this paper, the addresses of hosts in legacy edge network are considered transit addresses because they are globally reachable. The transit address of a host behind a classic address translator is its external address. The existence of classic address translators can thus be ignored without loss of generality.

2.2 Address Translation

To ensure that addresses are provider-independent inside an edge network, but globally routable outside, Six/One routers must translate *local* addresses in all packets that cross a border link: Destination addresses of incoming packets are always translated into edge addresses; source addresses of outgoing packets are always translated into transit addresses allocated by the provider via which the packets are sent. This enables edge networks to change providers without having to change their internal addressing. For packets exchanged with other upgraded edge networks, Six/One routers also translate between *remote* addresses. Address translations on either side of a packet exchange are then inverses of each other. This restores the end-to-end semantics of the edge addresses. To indicate to a remote Six/One router whether and how to translate a packet back into its original state, Six/One routers endow outgoing packet with a *Six/One extension header* including the packet's original source and destination addresses.

Figure 2 illustrates the operation of Six/One routers for a packet exchange between a local host and a remote host, which are both located in upgraded edge networks. The local host sends the first packet in the exchange, from its own edge address **`ABC::1`** to the remote host's edge address **`DEF::2`**. Neither of the Six/One routers has state for the packet exchange at the time it receives this first packet. The local Six/One router maps the local host's edge address onto transit address **`1000::1`**, and it performs mapping resolution to retrieve transit address **`2000::2`** for the remote edge address **`DEF::2`**. The local and remote edge addresses in the packet are then translated accordingly. The local Six/One router finally inserts a Six/One extension header into the packet to identify the original source and destination addresses. When the remote Six/One router receives the translated packet, it translates the addresses in the packet back to the original edge addresses, and then forwards the packet towards the remote host. Both Six/One routers cache the mapping between edge and transit addresses when handling the first packet of an exchange so that subsequent packets from the same exchange do not require renewed mapping resolution or mapping validation.

Six/One routers limit address translation to IP headers of packets; they do not attempt to find and translate addresses used in packet payloads. For packets exchanged with upgraded edge networks, the inverse translation by a remote Six/One router annuls potential address inconsistencies between the IP header and the payload in the translated packet. To avoid such inconsistencies also when the remote edge network is legacy, Six/One Router relies on application functionality for address translator traversal. Applications that reference addresses in packet payloads depend on this functionality already today, due to the wide deployment of address translators. It is hence safe to assume that those applications, which use addresses in packet payloads, also support address translator traversal.

2.3 Mapping Validation

Unsecured address indirection introduces a vulnerability to malicious packet diversion. An attacker capable of injecting bogus address mappings into indirection routers could direct packets for a particular edge address to a transit address of its choice. The objective of such an attack may be to eavesdrop on or modify the packets of a victim host from an arbitrary location in the Internet, or to flood a victim host or edge network with unwanted packets. In the former case, it is packets destined to the victim host that the attacker diverts, in the latter it is packets destined to the attacker itself.

Six/One Router relies on the security of the mapping resolution system to protect against malicious packet diversion. Six/One routers use a mapping if and only if it exists in the mapping resolution system. This rule applies naturally when a Six/One router resolves an unknown mapping. Six/One routers also consult the mapping resolution system to validate the mapping between source edge and transit addresses for a received packet with a Six/One extension header. The mapping is legitimate if it is included in the set of mappings returned by mapping resolution. Only then does the Six/One router translate the packet back into its original state. Otherwise, it drops the packet. In the example of figure 2, the local Six/One router performs mapping validation as part of resolving the mapping for destination address **`DEF::2`**, and the remote Six/One router performs mapping validation for source address **`ABC::1`** when it receives the packet. Since both Six/One routers cache the result of mapping resolution, the mappings for subsequent packets of the same packet exchange can be more efficiently validated without mapping resolution.

2.4 Domain Name Resolution

Since reachability at an edge address requires bilateral support of Six/One Router, hosts in upgraded edge networks must be reached at a transit address when contacted from legacy edge networks. This implies that DNS servers authoritative for upgraded edge networks must resolve domain names dependent on whether the querier is in an upgraded or in a legacy edge network. Queriers in upgraded edge networks must be provided with edge addresses, queriers in legacy edge networks with transit addresses.

For backwards compatibility, DNS servers must hence by default provide transit addresses. Special domain name resolution is then required only in the case where a host in an upgraded edge network resolves the domain name of a host in another upgraded edge network. Six/One Router uses two components for this:

- *New type of DNS resource records* – DNS servers authoritative for upgraded edge networks maintain resource records of a new type to store edge addresses for a given domain name. The standard A and AAAA resource records include the corresponding transit addresses.
- *DNS proxies* – Upgraded edge networks deploy DNS proxies via which local hosts transparently resolve domain names into edge addresses, if available, or otherwise transit addresses. Given a DNS query, DNS proxies initiate resolution processes for both edge and transit addresses. If edge addresses can be obtained, those are returned to the resolving host in standard A or AAAA resource records. Otherwise, retrieved A and AAAA resource records are returned.

Both DNS modifications affect only DNS servers in upgraded edge networks. This facilitates deployment. To limit resolution latency, DNS proxies can initiate the resolution of edge and transit addresses in parallel.

2.5 Multi-homing

An increasing number of edge networks multi-home to improve the performance and reliability of their Internet connectivity. Effective multi-homing therefore includes load balancing and fail-over between the providers of an edge network, which requires a mechanism to hand over both ingress and egress traffic between the providers. Six/One Router provides such a mechanism for packet exchanges between upgraded edge networks.

In Six/One Router, a provider handover for egress traffic occurs when the internal routing system of an edge network decides to redirect the outgoing part of a packet exchange via a new provider. The Six/One router adjacent to the new provider then translates local addresses in the packets to local transit addresses allocated by the new provider. It also performs mapping resolution to retrieve the remote transit addresses for the packets, and caches the result for future reference.

Load balancing and fail-over of ingress traffic requires feedback to remote edge networks about which transit addresses packets should preferably be sent to. Six/One Router uses for this purpose a Mapping Preferences message, which holds preference values for each of a set of transit address prefixes that all correspond to a given edge address prefix. Since edge addresses map one-to-one to the transit address from any particular provider, the edge and transit address prefixes in the Mapping Preferences message have the same length. This length can be arbitrary to facilitate preference feedback at variable granularity, ranging from an edge network's entire edge address prefix to a single edge address. A Mapping Preferences message can be returned to any source transit address from a received packet. It is intercepted and processed by the Six/One router that originated that packet.

To establish the validity of a received Mapping Preferences message, a Six/One router checks if its cache includes mappings that cover the mappings defined by the edge and transit address prefixes in the message. If this is so, the Six/One router tags those cache entries with the received preference values. This may cause it to select a new remote transit address for subsequent outgoing packets that are directed to the edge network from which the Mapping Preferences message was received. Mapping Preference messages that include mappings not in a Six/One router's cache are ignored because they cannot be validated.

3. ANALYSIS

This section analyzes the benefits of Six/One Router in terms of scalability improvements for the Internet routing system, and provider-independent addressing and multi-homing capabilities for edge networks. The section also evaluates the deployability of Six/One router in terms of backwards compatibility and incentives for early adoption.

3.1 Routing Scalability

Six/One Router improves the scalability of the Internet routing system by making the global routing table smaller and more stable, while still offering edge networks provider independence and the capability to redirect traffic between border links. Six/One Router achieves this through dynamic indirection of provider-independent edge addresses onto provider-allocated transit addresses. This makes the existence of edge addresses transparent to the global routing system, and hence enables their provider independence without impact on the Internet routing system. The global routing table thereby becomes smaller because routes to edge addresses are eliminated and routes to transit addresses can be more efficiently aggregated. The global routing table also becomes more stable because traffic can be redirected between an edge network's border links through a Mapping Preference message that is exchanged edge-to-edge, rather than through a global routing table update.

3.2 Re-addressing Avoidance

Six/One Router enables edge networks to change providers without changing the addresses of hosts or network entities internally. It accomplishes this based on exclusive use of provider-independent edge addresses inside an upgraded edge network. Provider-allocated transit addresses do not propagate into the edge network because they are translated into edge addresses in all inbound packets. All local addresses in use inside an upgraded edge network are therefore provider-independent and survive provider changes. What remains to be reconfigured in case of a provider change are the mappings for local edge addresses maintained by Six/One routers and the mapping resolution system.

3.3 Multi-homing

Six/One Router enables multi-homed edge networks to dynamically select the border link for incoming traffic from other upgraded edge networks in a way that is more scalable, more fine-granular, and, for re-selection in ongoing packet exchanges, more responsive than in today's Internet. It achieves this through direct mapping preferences announcements between Six/One routers in form of Mapping Preferences messages.

Border link selection is thus transparent to the global routing system. This increases routing scalability compared to today's Internet, where border link selection requires an update of the global routing table. Six/One Router spares the bandwidth and processing overhead for the global distribution of routing table update messages, and it avoids sub-optimal path selection in routers which routing table inconsistencies during a routing table update may provoke. Furthermore, border link selection can in both cases be accomplished per packet exchange. This is more fine-granular than in today's Internet, where a global routing table update typically affects a substantial fraction of an edge network's entire inbound traffic. Responsiveness of border link re-selection is a round-trip time between the Six/One routers handling the packet exchange. This is notably faster than the re-selection latency in today's Internet: Even global round-trip times are typically much less than 1 s in today's Internet [Co2005], whereas a global routing table update usually takes several minutes [La2000].

A border link in an ongoing packet exchange with a legacy edge network cannot be re-selected with Six/One Router. The local transit address is in this case in use by the remote host without indirection, so the packet exchange would break if that transit address would change.

3.4 Incremental Deployability

Six/One Router achieves incremental deployability through backwards compatibility and incentives for early adoption. Backward compatibility guarantees continued connectivity to the remaining legacy Internet. Incentives for early adoption in addition provide benefits for the deployers, even in the absence of notable Six/One Router deployment elsewhere. Thus, backwards compatibility and incentives for early adoption foster initial deployment, based on which further deployment can build.

Six/One Router provides backwards compatibility through support for unilateral address translation. This enables mutual reachability between upgraded and legacy edge networks. Six/One Router provides incentives for early adoption by simplifying the process of acquiring provider-independent edge addresses, especially for small edge networks. Classic provider-independent addresses are today typically allocated only to enterprise-size edge networks, in an attempt to contain negative impacts on global routing scalability. Address indirection may therefore for many edge networks be the only means for provider-independent addressing.

After an initial deployment basis has built up, use of Six/One Router may become attractive also to multi-homed edge networks with classic provider-independent addresses, due to the prospects of more responsive and more fine-granular traffic engineering on border links. The limitation of this being possible only for packet exchanges with other upgraded edge networks may at this point be acceptable given the existing deployment basis.

Eventually, remaining edge networks may consider deploying Six/One Router, even if they are neither interested in provider independence nor eager to pursue multi-homing themselves. One potential reason for this is increased reliability for packet exchanges with remote multi-homed edge networks. Another is to retain the end-to-end semantics of other edge network's edge addresses. The stability of remote edge addresses would make it easier to hardcode referrals to remote entities into local hosts or local network entities. Overall, these benefits may also help convincing edge networks to adopt Six/One Router for the benefit of a more scalable Internet routing system.

4. CONCLUSIONS

This paper proposes and evaluates a method for provider-independent addressing and efficient multi-homing in edge networks without compromising the scalability of the Internet routing system. This method is based on one-to-one translation between provider-independent edge addresses, for use inside edge networks, and aggregatable, provider-allocated transit addresses for routing in the Internet core. Address translation is backwards compatible because it can be pursued unilaterally. Only local addresses are then translated, and edge addresses are visible only in the edge network to which they belong. If deployed on both sides of a packet exchange, also remote addresses are translated. Translations on either side of the packet exchange are then inverses of each other, and edge addresses are visible end-to-end.

5. ACKNOWLEDGMENTS

The author would like to thank Jari Arkko, Pekka Nikander, Erik Nordmark, Marcelo Bagnulo, Iljitsch van Beijnum, Geoff Huston, Steven Blake, Lars Westberg, Mikko Särelä, Anders Eriksson, Börje Ohlman, Matthijs Mekking, William D. Herrin, Joel M. Halpern, Stéfane Bortzmeyer, Robin Whittle, Mark Doll, Lixia Zhang, Tony Li, Brian E. Carpenter, Lars Eggert, Pekka Savola, Magnus Westerlund, Jonne Soininen, Loa Andersson, David Ward, John Scudder, Gert Doering, Olaf Kolkman, Stig Venås, Thomas C. Schmidt, Kotikalapudi Sriram, Jordi Palet, András Császár, Jan M. Melen, Petri Jokela, Kristian Slavov, Patrik Salmela, Conny Larsson, and Attila Mihály for valuable and highly appreciated discussions, which have considerably shaped the design of Six/One Router.

6. REFERENCES

[Ao2007] Cedric Aoun, Elwyn B. Davies: Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status. Request for Comments 4966, July 2007.

[Be2007] Iljitsch van Beijnum: Modified Network Address Translation - Protocol Translation. Internet draft (work in progress), November 2007.

[Br2007] Scott Brim et al.: LISP-CONS: A Content distribution Overlay Network Service for LISP. Internet draft (work in progress), November 2007.

[Ca2007] Brian Carpenter: Shimmed IPv4/IPv6 Address Network Translation Interface (SHANTI). Internet draft (work in progress), November 2007.

[Ch2008] Noel Chiappa: Tunnel fragmentation/reassembly for RRG map-and-encaps architectures. IRTF Routing research group mailing list, on-line http://psg.com/lists/rrg/2008/msg00135.html, January 2008.

[Co2005] Les Cottrell: ICFA SCIC Network Monitoring Report. http://www.slac.stanford.edu/xorg/icfa/icfa-net-paper-jan05/, January 2005.

[Fa2007] Dino Farinacci et al.: Locator/ID Separation Protocol (LISP). Internet draft (work in progress), November 2007.

[Fa2007a] Dino Farinacci et al.: LISP Alternative Topology (LISP-ALT). Internet draft (work in progress), November 2007.

[Fu2007] Vince Fuller: Scaling Issues with Routing and Multi-homing. On-line http://www.vaf.net/~vaf/apricot-plenary.pdf, February 2007.

[Ha2005] Tony Hain: A Pragmatic Report on IPv4 Address Space Consumption. Internet Protocol Journal, volume 8, number 3, September 2005.

[Hu2003] Geoff Huston: IPv4: How long do we have? Internet Protocol Journal, volume 6, number 4, December 2003.

[Hu2008] Geoff Huston: BGP Table Size. On-line http://bgp.potaroo.net/cidr/, February 2008.

[Hu2008a] Geoff Huston: IPv6 / IPv4 Comparative Statistics. On-line http://bgp.potaroo.net/v6/v6rpt.html, February 2008.

[Jen2007] Dan Jen et al.: APT: A Practical Transit Mapping Service. Internet draft (work in progress), November 2007.

[La2000] Craig Labovitz et al.: Delayed Internet Routing Convergence. Proceedings of ACM SIGCOMM, August 2000.

[Le2008] Eliot Lear: NERD: A Not-so-novel EID to RLOC Database. Internet draft (work in progress), January 2008.

[Lw2007] Darrel Lewis et al.: Interworking LISP with IPv4 and IPv6. Internet draft (work in progress), December 2007.

[Li2007] Tony Li, ed.: Design Goals for Scalable Internet Routing. Internet draft (work in progress), July 2007.

[No2000] Eric Nordmark: Stateless IP/ICMP Translation Algorithm. Request for Comments 2765, February 2000.

[RAWS] David Meyer et al.: Report from the IAB Workshop on Routing and Addressing. Request for Comments 4984, September 2007.

[RRG] IRTF Routing Research Group (RRG). On-line http://www3.tools.ietf.org/group/irtf/trac/wiki/RoutingResearchGroup, February 2008.

[Sr2001] Pyda Srisuresh, Kjeld Borch Egevang: Traditional IP Network Address Translator (Traditional NAT). Request for Comments 3022, January 2001.

[Ts2000] George Tsirtsis, Pyda Srisuresh: Network Address Translation - Protocol Translation (NAT-PT). Request for Comments 2766, February 2000.

[Vo2007a] Christian Vogt: Six/One: A Solution for Routing and Addressing in IPv6. Internet draft (work in progress), November 2007.

[Vo2008b] Christian Vogt: DNS Map - A DNS-Based Resolution System for IP Address Mappings. On-line http://users.piuha.net/chvogt/pub/2008/vogt-2008-dns-map.pdf, February 2008.