canarie

CANARIE Network Routing Policy

Version 1.0

25 August 2010

1 Introduction

1.1 Purpose of Document

This document outlines the use of unicast routing protocols in the CANARIE IP backbone and the mechanisms used to influence network traffic flow in order to meet the routing policy objectives. The contents of this document equally apply to the IPv4 and IPv6 protocols, except where explicitly mentioned otherwise. Multicast routing protocols are covered in a separate document.

This document is written for operational and engineering personnel of the CANARIE Network, GigaPoP operators, international network peers, and international network exchange point operators. This document will be updated as operational requirements evolve.

1.2 Overview of CANARIE Network

CANARIE Network is a high performance hybrid network infrastructure to support Canadian research and higher education communities. This infrastructure enables CANARIE to offer the tradition IP network services and Lightpath (an end-to-end connection) services to users.

The SONET infrastructure is an important layer of the CANARIE network. It is built using a mixture of CANARIE-lit wavelengths, carrier-leased wavelengths and wavelengths swapped with like-minded organizations. All wavelengths are 10Gbps unprotected, point-to-point connections, which terminate on transport switches located in CANARIE Network Points of Presence (PoPs).

The CANARIE IP network, is built based on Lightpaths at, or a fraction of, 10Gbps speed, linking five high performance routers across Canada, provides advanced IP services to Canadian R&E community,

Institutions access CANARIE IP service through aggregation points called GigaPoPs. The GigaPoPs are operated by provincial Research and Education (R&E)networks (RAN) and connect to CANARIE backbone routers either directly, i.e. router to router, or through the CANARIE Network infrastructure. These connections operate at 1Gbps or 10Gbps.

The CANARIE IP network is used to carry traffic between R&E institutions only. As access to commodity IPv6 Internet is increasingly important for the R&E community, CANARIE has extended its IPv6 service to offer full commodity IPv6 routing to the GigaPoPs. Commodity IPv6 Internet connectivity is provided through peering with a number of commercial ISPs, by way of direct links or through Internet eXchange Points (IXPs).

CANARIE is responsible for network engineering as well as the day-to-day operation of the network, performed by the Network Operations Centre (NOC) within CANARIE.

1.3 Routing policy objectives

The CANARIE Network layer 3 service routing policy seeks to accomplish the following:

- enforce CANARIE’s Acceptable Use Policy (AUP)
- minimize path latency
- accommodate the requirement for route diversity
- enforce symmetric routing

1.3.1 AUP compliance

CANARIE's AUP is an "institutional AUP". Institutions that are given approval to access the CANARIE network may send and receive traffic regardless of traffic type. Institutions can only connect to the CANARIE network through CANARIE approved GigaPoPs. In addition, all approved institution routes must be registered in the CANARIE Routing Registry (CRR) in order for their routes to be accepted and advertised over the CANARIE IP backbone.

1.3.2 Minimum path latency

Minimum path latency between any two connected endpoints is desirable for seamless operation of real-time applications.

Path latency is the sum of transmission delay, processing and queuing delay, and propagation delay. In an uncongested long haul IP network, transmission as well as processing and queuing delays, is negligible relative to propagation delay. Propagation delay is determined by the physical link span. In order for the routing protocols to choose the lowest latency path, the metrics assigned to each backbone link must be representative of the physical distances or measured latencies between the core routers.

1.3.3 Route diversity

Reliability and availability are important characteristics of any operational network. Route diversity has direct impact on network resilience and is therefore necessary to minimize the possibility of network segmentation in the face of network component failure. Route diversity is also required to support traffic engineering.

Route diversity and minimum path latency are orthogonal requirements. A route diversity scenario will therefore incorporate a lower and a higher latency path.

1.3.4 Symmetric routing

The performance of some network services and applications can be adversely affected by asymmetric routing. As well, firewalls do not normally permit asymmetrically routed connections to be established. To minimize application breakage and for ease of debugging routing problems, consistent symmetric routing is desired.

2 CANARIE IP network

2.1 Topology

The present topology of the core CANARIE IP network is depicted in Figure 1. It is important to note that for the commodity Internet IPv6 service, some dedicated links and a router have been added, resulting in slightly different topologies for IPv4 and IPv6. IPv6 only elements are drawn using transparent lines and shapes.

The common IPv4/IPv6 topology of the CANARIE core network is comprised of five routing nodes, seven internal and five external network segments. Each segment is a lightpath that operates at, or a fraction of, 10Gbp. The nodes are located in Calgary, Winnipeg, Toronto, Montréal, and Halifax. Of the seven internal segments, four link the routers located in adjacent cities, while the remaining three link the routers located two cities away.

Five external network segments connect the CANARIE network to the following three R&E Internet exchanges: Pacific Wave in Seattle, StarLight in Chicago, and MANLAN in New York.

IPv6 specific elements of the topology are a node in Vancouver, one additional internal segment (Vancouver – Calgary) and four additional external segments, one to commercial Internet exchange in Seattle and one in Toronto and two direct links to ISP Tata Communications in Vancouver and Montréal.

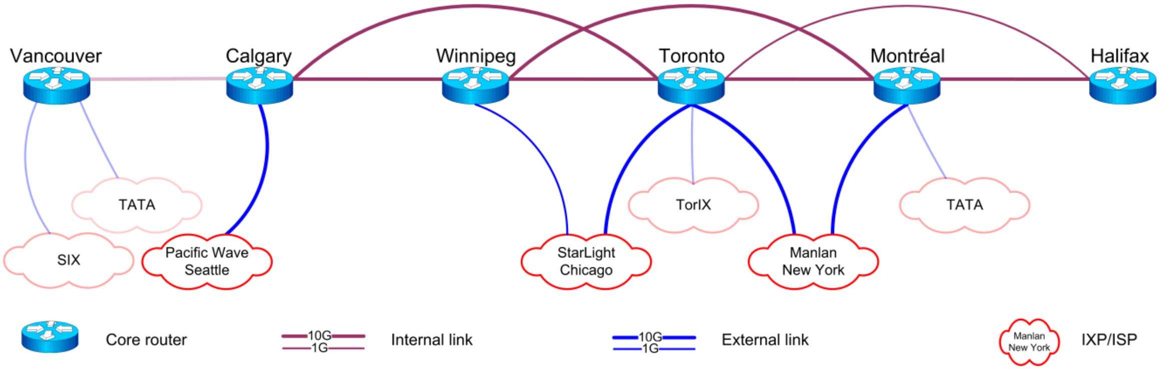

Figure 1: CANARIE IP internal and external network segments

In order to accommodate the routing objectives of minimum latency and route diversity in the design, a balance was to be obtained by choosing the lowest latency paths for the 3 segment linking nodes in non-adjacent cities. The four segment linking nodes of adjacent cities, on the other hand, would sacrifice some latency performance in order to provide route diversity. In this manner, the effective operating range of real-time applications, sensitive to latency, would not be compromised.

3 Routing Protocols

3.1 Interior Gateway routing Protocol

An Interior Gateway Routing Protocol (IGP) is required to run between the routers in order to create and maintain an up-to-date topological database of the IP backbone network. This topological database is required for calculation of the shortest paths between nodes and forms the basis for keeping the internal Border Gateway Routing Protocol (iBGP) peering sessions up between the routers.

CANARIE network IGP uses the Intermediate System-to-Intermediate System (IS-IS) routing protocol with a single IS-IS Level 2 area defined.

3.1.1 IS-IS Metrics

IS-IS backbone metrics are based on segment latency. Usage of “wide” IS-IS metrics allows direct mapping of latencies (□S to routing protocol metrics. The baseline network segment latencies are shown in Table 1.

network segment	**baseline latency in □a**
Calgary - Vancouver	5500
Calgary - Winnipeg	10950
Calgary - Toronto	20250
Winnipeg - Toronto	14830

Winnipeg - Montréal	15760
Toronto - Montréal	11717
Toronto - Halifax	12612
Montréal - Halifax	10415

Table 1: CANARIE IP network segment baseline latencies

3.1.2 Route Selection

The chosen network topology and IS-IS metric assignment makes route selection trivial. A minimum of two paths exists between every router of which there can only be one single best path. If any part of the primary path fails, a backup path is chosen.

3.2 External Gateway Protocol (EGP)

CANARIE IP network uses BGP as the EGP. iBGP is used to exchange external destination reachability information within the Autonomous System (AS). The iBGP peering configuration is full mesh; meaning that each router maintains N-1 iBGP peering. iBGP peering is configured using a globally unique loopback interface IP address. iBGP session establishment depends on IS-IS path selection.

GigaPoPs, RANs and international networks maintain an eBGP peering session with the core router(s) they are connected to. The BGP session is always configured using IP addresses of the directly connected interface. The peering configuration, route acceptance and announcement, are based on the routing policy and objects registered in the CANARIE Routing Registry (whois.canarie.ca).

3.2.1 BGP Route Acceptance

3.2.1.1 Network Peer Prefix Filtering

All routes announced by GigaPoPs are subject to prefix filtering prior to being added to the CANARIE IP network routing tables. The prefix filter list for each GigaPoP is generated from the route data found in the CRR and updated daily.

International peer routes are not systematically prefix filtered, although, where supporting mechanisms are in place to do so, it is the preferred approach.

3.2.1.2 IPv4 Bogon Prefix Filtering

IPv4 bogon route filtering is applied to all international IPv4 peering sessions where prefix filtering is not implemented. Bogon prefix filter combines Martians (as defined by RFC 1918 and RFC 5735) and network blocks that have not been allocated to regional Internet registries by IANA. BGP update messages for these prefixes are ignored.

Bogon filter is generated based on the ‘fltr-bogons’ filter-set object registered in both RADb (whois.radb.net) and RIPE (whois.ripe.net) Internet Routing Registries.

3.2.1.3 IPv6 prefix filtering

IPv6 generic route filtering is applied to all international IPv6 peering sessions where prefix filtering is not implemented. The filter accepts routes up to /48 from the known IANA allocations to RIRs. It rejects well-known IPv6 address blocks, as a multicast range (RFC3513), a 6bone or a 6to4 range of addresses.
IPv6 route filter is generated based on the ‘fltr-v6’ filter-set object registered in CANARIE routing registry. This object is generated and maintained by the CANARIE NOC.

3.2.1.4 AS Path filtering

As-path filtering is not performed.

3.2.1.5 Community-based filtering

The tagging of peers’ routes with community attributes is not required for route acceptance, except for temporary project routes as defined in section 3.2.3.5 of this document.

3.2.2 BGP Route selection

3.2.2.1 BGP Local Preference Attribute

The BGP Local Preference (LP) Policy is an arbitrary but logical policy used to make egress routing decisions between routes heard from more than one peer Autonomous System (AS).

For multi-homed peers, configuring the same LP value on each of the multiple peerings permits load splitting between multiple paths. Choice of the active path, or paths, will depend on other factors such as IGP metrics and the use of MEDs.

BGP LP attribute values are assigned on an AS basis, irrespective of the homing scenario. CANARIE GigaPoP routes will be assigned the highest LP value. For international peers, in general, LP will be assigned in a manner which inversely reflects the degree to which the AS acts as a transit network. This is to ensure that direct connections with non-transit providing peers are utilized.

LP value assignment guidelines are listed in table 2 below; however final assignment is determined on a case by case basis after careful consideration of traffic flow characteristics and network path performance.

An up to date listing of the current CANARIE BGP peering sessions and associated LP attributes can be found in CANARIE Routing Registry.

Network scale	**BGP Local Preference attribute**
CANARIE GigaPoP	500 (510)
National R&E Network (NRN)	400
NRN aggregator network	300
ITN service provider	200
Commodity ISP (IPv6 only)	100

Table 2: LP value assignment rule of thumb

3.2.2.2 MULTI_EXIT_DISC (MED) Attribute

Advertisement of MEDs by a CANARIE multi-homed peer, and their acceptance, will influence the egress route selection such that the preferred path back into the peer network will be chosen based on the lower value of this BGP attribute. If the MED attribute is based on internal metrics representing segment latency, this will result in CANARIE Network traffic to egress at the peering point closest to the destination. See figure 2.

CANARIE will accept MEDs from multi-homed peers where, in general, it will result in lower latency paths between end networks. In order to make this decision, a study of the peer network topology, routing policy, and traffic flow characteristics may be required.

CANARIE will not advertise MEDs to multi-homed peer networks from which it accepts MEDs, in order to avoid asymmetric routing.

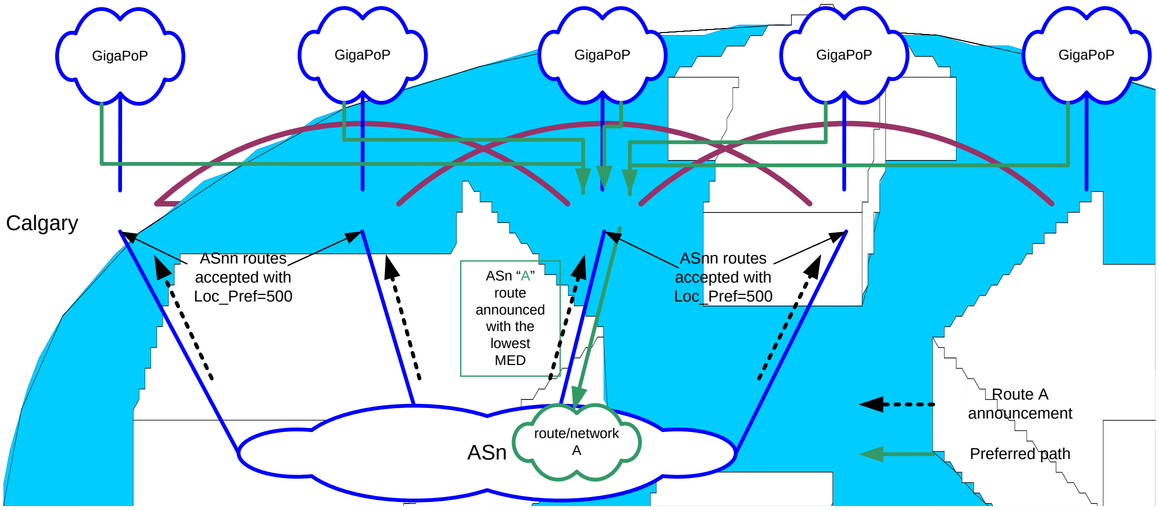
GigaPoP
GigaPoP
GigaPoP
GigaPoP
GigaPoP
Calgary
ASnn routes accepted with Loc_Pref=500
ASn "A" route announced with the lowest MED
ASnn routes accepted with Loc_Pref=500
Route A announcement
Preferred path
ASn
route/network A

Figure 2: Use of MED to influence CANARIE traffic egress

3.2.3 Route tagging (BGP Communities)

3.2.3.1 General

Route tagging is a flexible mechanism for implementation of routing policy within an AS.

RFC 1997 defines a BGP community as a group of destinations that share some common property.

In accordance with the RFC, the first two octets of this 32 bit attribute will be the AS number of the AS whose policy is being implemented by the use of the tag. The last two octets are associated with a policy implementation. Tag notation in this document will follow the format convention of AS#:number.

The BGP specifications have been extended to support four-octet AS numbers. Obviously, the format convention described above would not work for four-octet AS numbers. The four-octet AS

Specific Extended Communities, defined in RFC5668, would be required instead. Because of a limited software support by router vendors and no need for their practical use, the extended communities are currently not utilized.

6509:x tags received from international R&E network peers will be ignored and cleared upon reception of the BGP announcement.

3.2.3.2 Entry point tagging

In order to facilitate debugging of routing, all accepted routes are tagged with a BGP community identifying the entry point of the route into CANARIE Network. The tags will be transitive in order for GigaPoP Operators and international network peers to also quickly identify the route entry point. Table 3 lists the entry point tags. For example, a route received by the router in Montréal will be tagged with 6509:65040 before being propagated.

6509:65010	Calgary
6509:65020	Winnipeg
6509:65030	Toronto
6509:65040	Montréal
6509:65050	Halifax
6509:65060	Vancouver

Table 3: Entry point tags

3.2.3.3 ITN tagging

In addition to the "point-of-entry" tagging, International Transit Network (ITN) Service tags are used to simplify the administration and coordination of international transit across Internet2, CANARIE, and STARTAP. ITN tags will be transitive. Table 4 lists ITN tags used in CANARIE Network.

6509:2500	do not transit
6509:2501	transit to STARTAP
6509:2502	transit to STARTAP and Internet2

Table 4: ITN tags

3.2.3.4 Custom international transit tagging

International network peers wishing limited transit across CANARIE Network to one or more AS, may do so by requesting the service of the CANARIE NOC. If the second party to the transit request is not an ITN participant, an explicit agreement will be required, and should be arranged either directly or through the CANARIE NOC. Once an implicit or explicit bilateral agreement is reached, each party's routes will be tagged with 6509: destinationAS# in order to enable the transit.

3.2.3.5 Commodity Internet IPv6 routes tagging

IPv6 routes accepted from commercial network peers will be tagged with a BGP community as listed in Table 5. The tag will be transitive and propagated to GigaPoPs, to allow GigaPoP Operators to easily identify and apply policies to these routes.

6509:64600	Commercial IPv6 routes

Table 5: commodity IPv6 routes

3.2.3.6 Special project tagging

In principle, non approved institutions can also access the CANARIE Network on a meritorious project basis. Project basis implies two things: the project is defined in time and involves two, or in rare cases more than two, end points.

GigaPoP Operators wishing to propagate special project routes must, in addition to the normal CRR registration process, tag the route with 6509:destinationAS#.

3.2.3.7 GigaPoP Operator LP tweaking

By default, Canadian GigaPoP Operator routes will be given a LP of 500. However, for scenarios where an end institution is multi-homed to two different GigaPoPs, or where a GigaPoP Operator receives all or a subset of CANARIE routes directly from another GigaPoP, the ability to increase the local preference may be of value to influence traffic flow. Table 6 lists the GigaPoP LP tweaking tags. Section 6.1 describes a scenario where this service may be useful.

6509:65510	Increase LP to 510

Table 6: GigaPoP LP tweaking tags

3.2.4 Route Announcement

3.2.4.1 AS_Path Prepending

AS path prepending is not currently used in CANARIE Network.

3.2.4.2 MEDs

In order to obtain lowest latency symmetric routed paths, when used, MEDs will be based on IGP metrics.

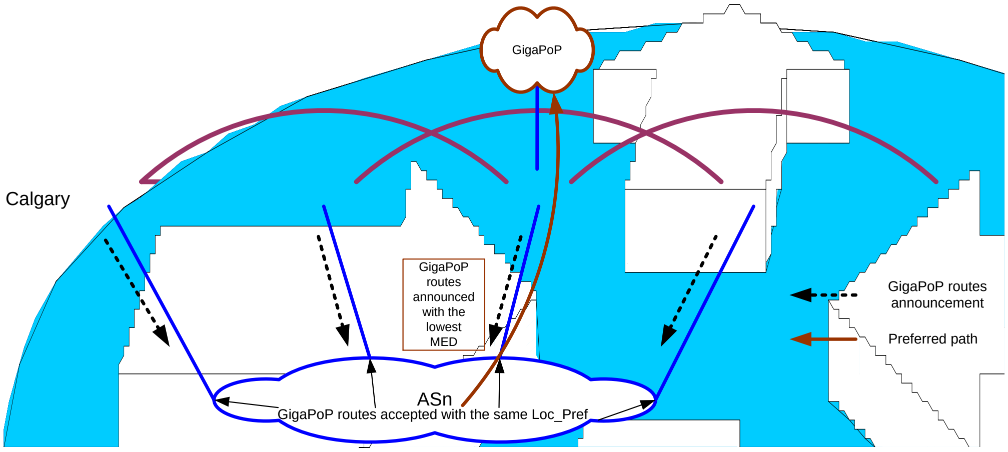
GigaPoP
Calgary
GigaPoP routes announced with the lowest MED
ASn
GigaPoP routes accepted with the same Loc_Pref
GigaPoP routes announcement
Preferred path

Figure 3: Use of MEDs to influence CANARIE Network traffic ingress

As can be seen from figure 3, the CANARIE Network multi-homes to the StarLight and MAN LAN international R&E exchanges from two geographically distinct points. The majority of international network peers with which CANARIE peers through those exchanges are physically present at the exchanges. Due to this topological situation, it generally makes sense that CANARIE announce MEDs to those international peers.

3.2.4.3 Route announcements

CANARIE will advertise all CANARIE, GigaPoP, member and international network peer routes to the GigaPoPs.

CANARIE, GigaPoPs and members' routes will be advertised to international network peers, with the exception of routes for which international network peers have requested ITN or custom transit service across CANARIE's Network (refer back to section 3.2.3 for more information on these services).

3.2.5 BGP Router Configuration

3.2.5.1 Soft BGP Reconfiguration

Soft BGP reconfiguration is enabled on the CANARIE backbone routers. This feature allows policies to be changed and implemented without tearing down the BGP session, thus contributing to the stability of the overall routing system.

3.2.5.2 BGP Dampening

BGP route dampening is explicitly disabled on the CANARIE backbone routers.

3.2.5.3 BGP Authentication

BGP authentication is configured on a case by case basis and subject to a bilateral agreement with a network peer.

4 Packet filtering

4.1 Unicast Reverse Path Forwarding (RPF) filtering

To help ensure symmetric routing across the CANARIE Network, GigaPoP ingress interfaces are configured with unicast reverse path forwarding (RPF) checking enabled. Contravening packets are dropped.

Although this is a packet filtering technique, it is mentioned in this document because the packet filtering decision relies on routing table information. RPF check is also a security mechanism that limits source address spoofing.

5 Scenario examples

5.1 Multi-homed End-User Institution

Scenario: A CANARIE Network end-user institution is multi-homed to two independent GigaPoPs. It may wish to prefer one path and use the other as backup only. See figure 4.

Policy implementation:

1) Both GigaPoP Operators register the end-user institution's route(s) in the CRR in order to allow proper configuration of prefix-filters.
2) The GigaPoP Operator on the preferred path announces the route(s) to CANARIE with tag 6509:65510. The other GigaPoP Operator announces the route(s) without tags.
3) CANARIE Network router tags the route(s) according to entry point.
4) Routes tagged with 6509:65510 are given a LP of 510. The same untagged routes are given a LP of 500.
5) Packets entering the CANARIE Network are checked for RPF. Non-conforming packets are dropped.

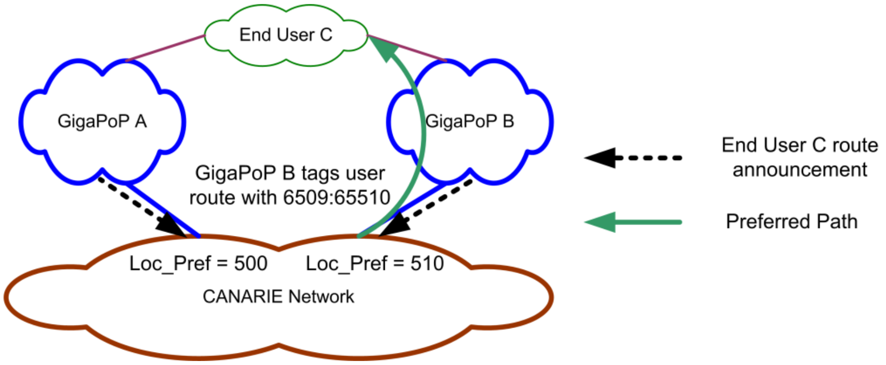

Figure 4: Use of tagging by GigaPoP Operator to influence CANARIE Network LP.

5.2 GigaPoP-GigaPoP backup

Scenario: Two GigaPoPs with direct peering wish to provide each other with a backup route to CANARIE Network. See figure 5.

Policy implementation:

1) Both GigaPoP Operators register their end-user institution's routes in the CRR.
2) In addition, both GigaPoP Operators need to express their routing policy in the CRR that reflects this routing scenario.
3) The combination of actions 1) and 2) will permit proper configuration of prefix-filters on CANARIE routers.
4) CANARIE router tags the route(s) according to entry point.
5) Backup path is used in case of GigaPoP local loop break.
6) Packets entering CANARIE Network are checked for RPF. Non-conforming packets are dropped.

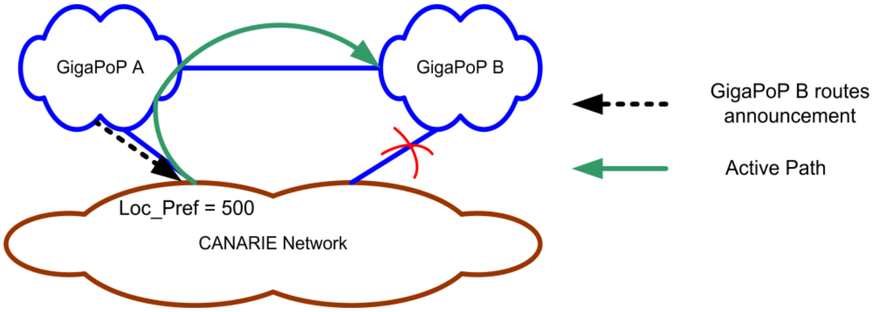

Figure 5: Directly connected GigaPoPs providing mutual backup to CANARIE Network.

5.3 "Triangle" Multi-homing of International Peer

Scenario: An international R&E network peer node is multi-homed to two geographically separate CANARIE nodes (normally the case for peerings through the StarLight and MAN LAN layer 2 exchanges). See figure 6.

Policy implementation:

1) If an easy mechanism is available, e.g. IRR, then prefix filtering is enforced. Otherwise int'l peer route announcements must pass the IPv4 bogon and IPv6 prefix filter tests.
2) CANARIE router tags the routes according to their entry point.
3) CANARIE router tags the routes with appropriate ITN Service tag.
4) CANARIE router announces routes with MEDs.
5) Packets entering CANARIE Network are checked for RPF. Non-conforming packets are dropped.

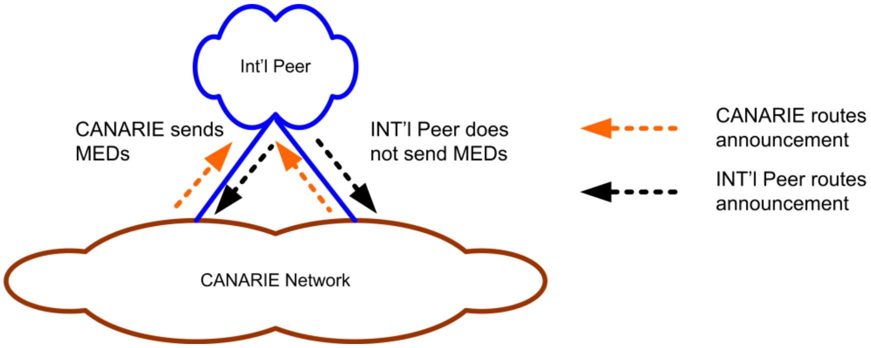
Int'l Peer
CANARIE sends MEDs
INT'I Peer does not send MEDs
CANARIE Network
CANARIE routes announcement
INT'I Peer routes announcement

Figure 6: International peer multi-homed in "triangle" topology.

5.4 Custom International Transit Service

Scenario: Two international R&E network peers wish transit across the CANARIE Network to reach each other but do not wish full ITN service. See figure 7.

Policy implementation:

1) The CANARIE NOC receives the transit request and verifies implicit or explicit bilateral agreement.
2) The CANARIE NOC tags the routes of the two international R&E networks wishing transit with 6509:AS#, where AS# is the AS of the R&E network they which to transit to.

3) International routes must pass the IPv4 bogon and IPv6 prefix filter tests.
4) CANARIE router tags the routes according to their entry point.
5) Packets entering CANARIE Network are checked for RPF. Non-conforming packets are dropped.

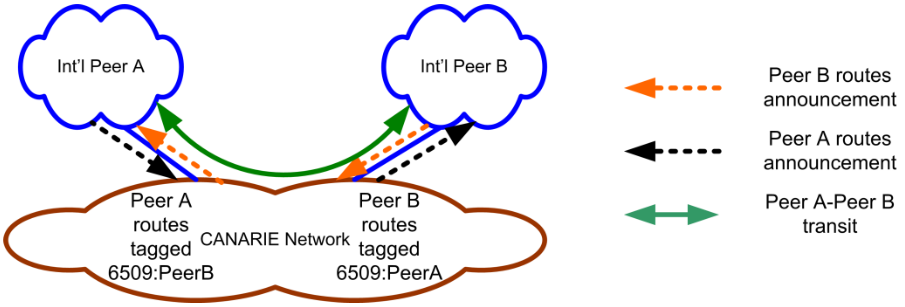
Int'l Peer A
Int'l Peer B
Peer A routes tagged 6509:PeerB
CANARIE Network
Peer B routes tagged 6509:PeerA
Peer B routes announcement
Peer A routes announcement
Peer A-Peer B transit

Figure 7: Custom international transit service