LHCONE BGP Communities

Mike O'Connor, Network Engineer

ESnet Engineering Group

U.S. DEPARTMENT OF **ENERGY**
Office of Science

BERKELEY LAB

BGP Communities Overview

The BGP communities path attribute is an optional transitive attribute of variable length.

Prior to the creation of BGP communities, the distribution of routing information was solely based on either IP address prefixes or on the value of the AS_PATH attribute.

BGP communities facilitate and simplify the control of routing information by grouping destination prefixes so that the routing decision can also be based on the identity of a group, significantly simplifying a BGP speaker's configuration that controls distribution of routing information. (RFC-1997)

BGP Communities Overview

External policy control - BGP communities were devised as a way to control routing policy in an upstream provider network. This is the primary reason for considering the use of communities in the LHCONE context.

Internal policy control - Communities that are assigned and consumed within an NSP network to control BGP route updates to customers, peers and upstream transit providers.

Communities are assigned and policy is implemented in the BGP import and export policy chains on a peer by peer basis.

While External policy control is the primary area of interest in using BGP communities within the LHCONE context, Internal policy control employing communities is a very sound and useful approach for VRF overlay networks.

Well Known Communities

There exists a small set of well known communities that should be implemented when an AS decides to implement community based policy. These well known communities are applied to route prefixes and affect policy in the upstream or peer network.

NO_EXPORT (0xFFFFFF01) – distribute within the upstream AS or AS confederation, do not export external to it ie:EBGP.

NO_ADVERTISE (0xFFFFFF02) – apply to the directly connected edge router, but do not advertise within the AS ie:IBGP.

NO_EXPORT_SUBCONFED (0xFFFFFF03) – Do not export external to the AS or across any existing AS confederation boundary.

The remaining community attribute values shall be encoded using an autonomous system number in the first two octets. The semantics of the final two octets may be defined by the autonomous system.
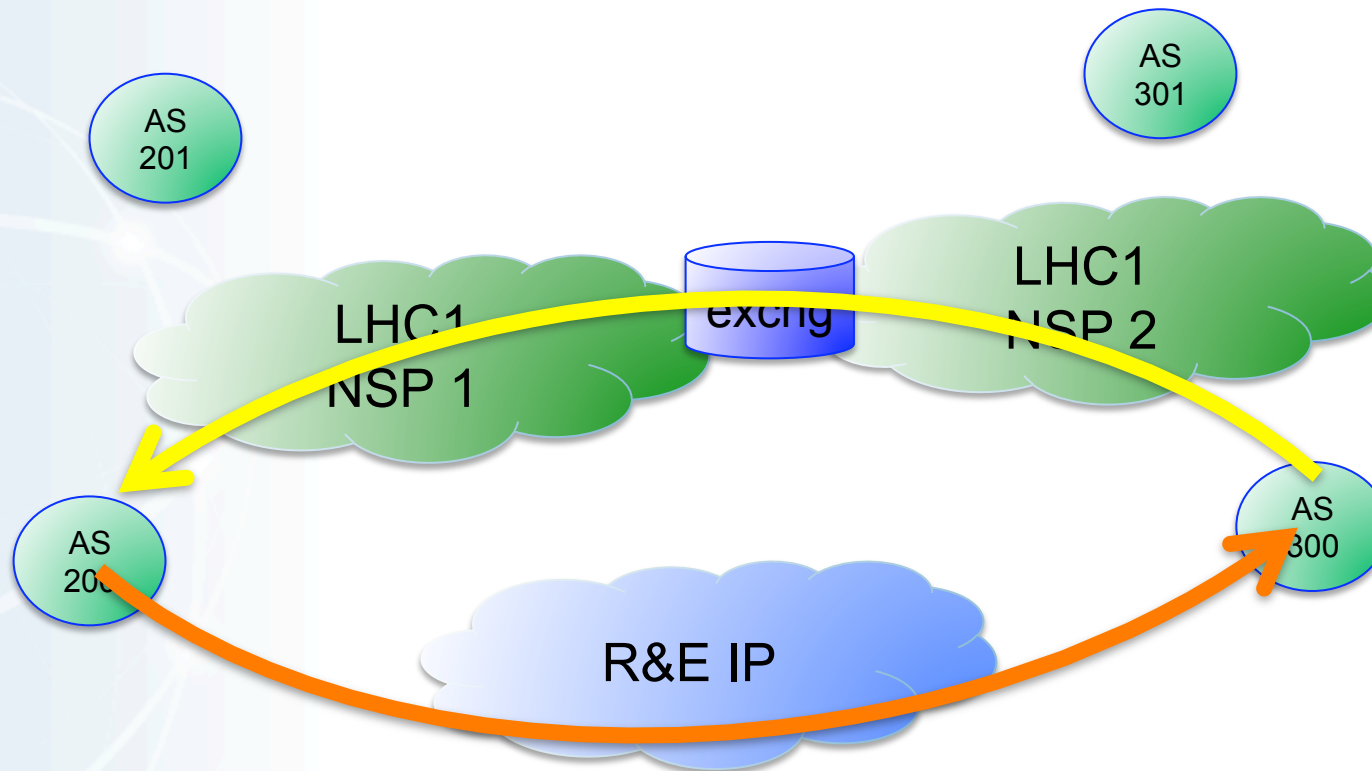
BGP Route Policy Without Communities
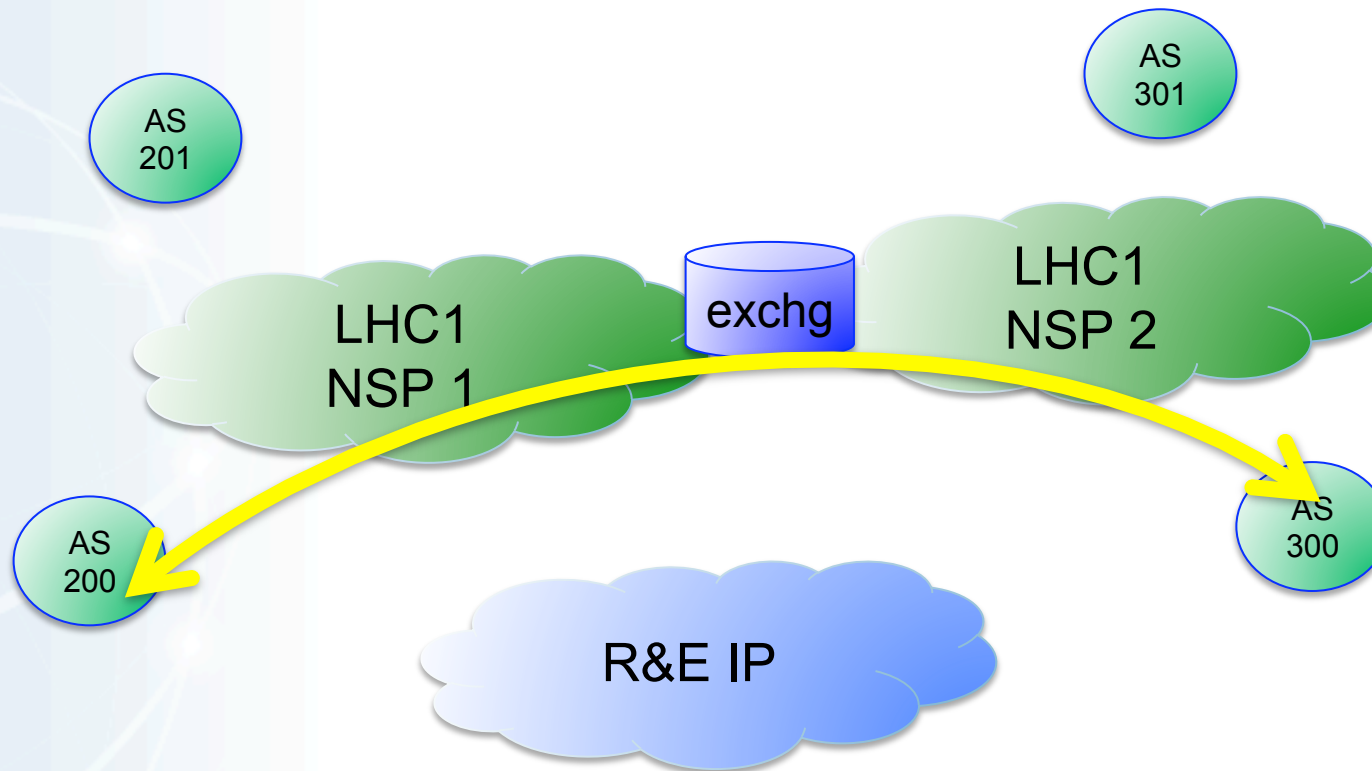
AS 200 prefers to connect to AS 300 via R&E IP, so they deny AS 300 prefixes in route updates from LHCONE in their import policy.

BGP Route Policy Without Communities
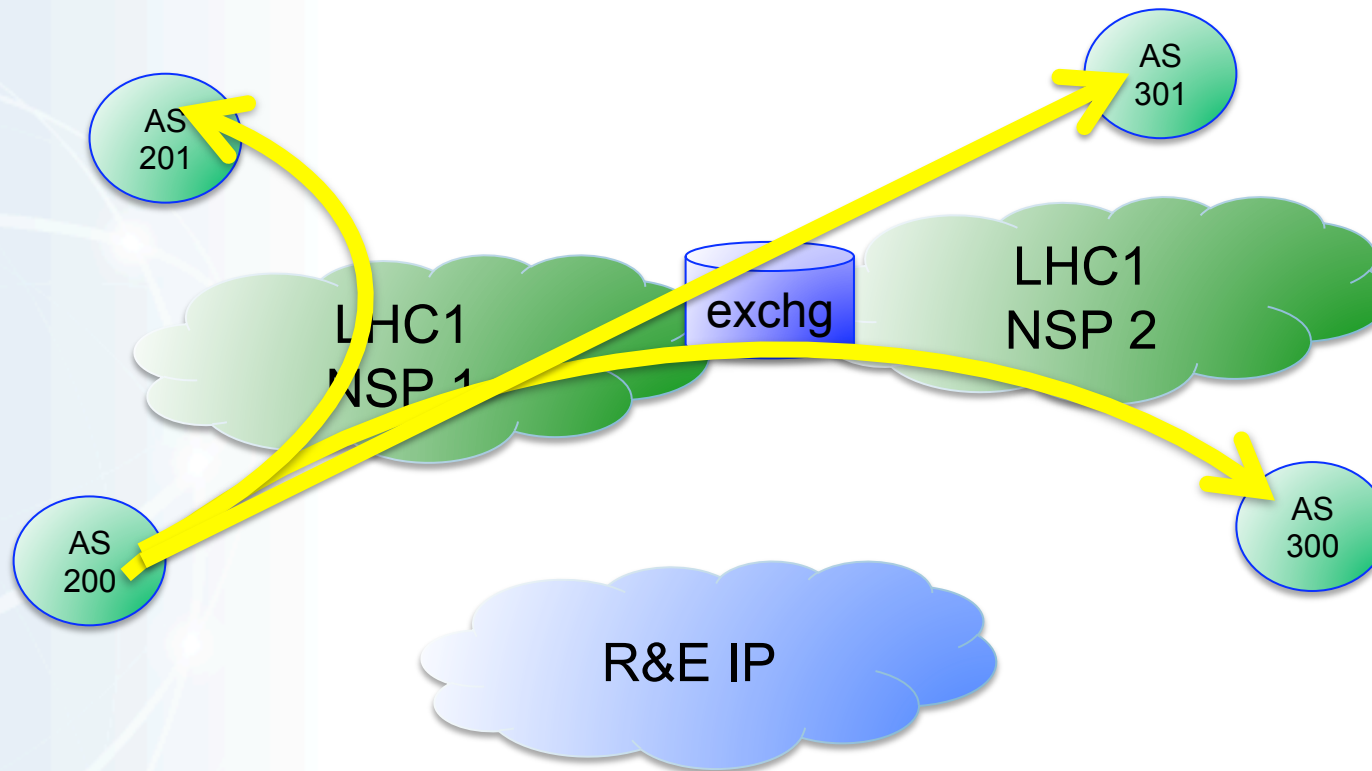
Without communities, AS 200 can not alter their BGP export policy without affecting connectivity with AS 201 and AS 301.

BGP Route Policy Without Communities

AS 200 configures DENY AS 300 in their LHCONE import policy chain, causing asymmetric routing. AS 300 continues to forward into LHCONE to reach AS 200.

Communities to control route advertisements in LHCONE

Scaling the community list to accommodate growth will require a convention for generating meaningful communities asynchronously and out of band, without having to maintain and distribute lists that are very difficult to maintain uniformly. We would need a formula.

ie: ASN:NO_EXPORT

Each NSP would be required to configure a complete set of LHCONE communities, in addition to a unique policy chain for each of it's individual customer ASNs.

A customer site would need to configure a complete set of LHCONE communities for as many remote collaborator networks as it intends to deprecate. Prefixes may be tagged with multiple communities but the limit is relatively low and inadequate for this scheme.

BGP Route Policy
With Communities
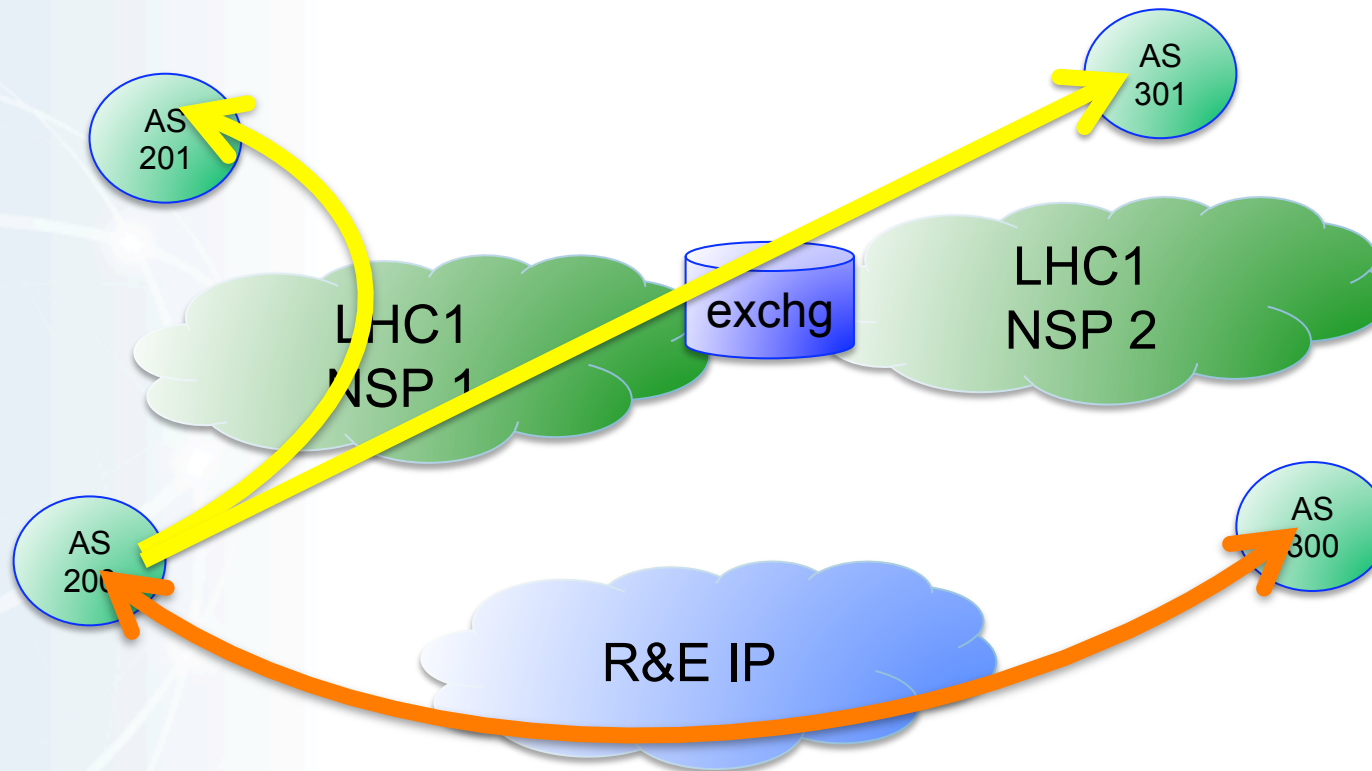
AS 200 tags **300:NO_EXPORT** on all of it's prefixes and NSP 2 policy will filter it from the LHCONE route update to AS 300, assuming the LHCONE is uniformly configured to accept and adjust policy based on community tags.

Alternative Approach

The previous example illustrates how the uniform implementation of an LHCONE BGP community policy across all NSPs allows a single research institute to control their LHCONE flows to a given remote ASN in **both directions** with an **in-band control protocol**.

Alternatively, an **out of band method** might take the form of an email to the network admin of the opposite ASN requesting that they deny your prefixes in their LHCONE BGP import policy. While this doesn't scale well, neither does the in-band method.

Internal BGP Policy Control with Communities

A common and convenient method for setting MED values is to have the edge router assign MEDs based on an IGP metric. Unfortunately VRF overlay network implementations don't support this functionality and another MED assignment scenario must be implemented.

ESnet assigns a geographic MED to all route prefixes it receives from it's LHCONE customers. Each NSP peering decrements the MED in the export policy if the prefix regional community matches the edge region of the NSP peering then the MED for that prefix is decremented.

Internal BGP Policy Control with Communities

For example:

ESnet assigns a default MED of 100 to each prefix received from our LHCONE participating customers and tags those same prefixes with a regional community.

The ESnet BGP export policy in the Chicago region will decrement the MED for FNAL prefixes by 10 upon export at the Starlight exchange, making this the preferred path to FNAL over LHCONE for NSPs peer with ESnet at Starlight. Similarly BNL will have it's MED decremented upon export at MANLAN.

Conclusions

- **No distribution protocol**. BGP communities must be manually configured on that router, this extends to the required policy chain as well. This type of high maintenance configuration is manpower intensive, slow and error prone, especially when exacting compliance is required across the entire LHCONE community.

- The logical inverse of the example policy presented, **deny all with explicit accept** doesn't scale at all and is not an option. The size of the LHCONE community is greater than community tag limits will accommodate.

- All NSPs would have to agree to fully participate, failure to do so would create asymmetric routing as a participating site implements a deny policy for another site and the community tag fails to prevent export at the far end.

- There are LHCONE sites that don't have their own ASN and would not be able to participate at all in this scheme.

Even if fully implemented, the scheme does not scale beyond several denied sites per participant, severely diminishing the return on the investment in manpower, increased complexity and coordination.

Questions?

Michael O'Connor

ESnet Network Engineer

moc@es.net

631 344-7410

ESnet Template Examples

Lawrence Berkeley National Laboratory

U.S. Department of Energy | Office of Science