

---

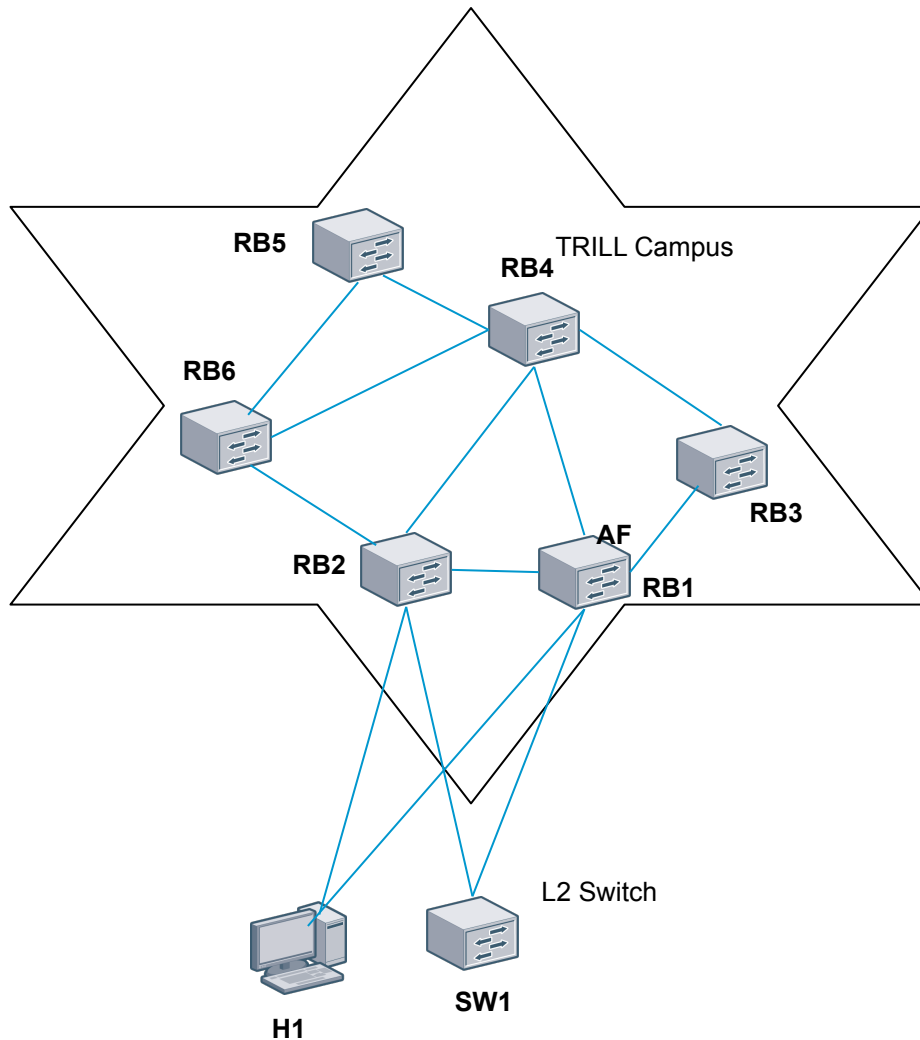
# TRILL Multihoming Proposal



Pathangi Janardhanan

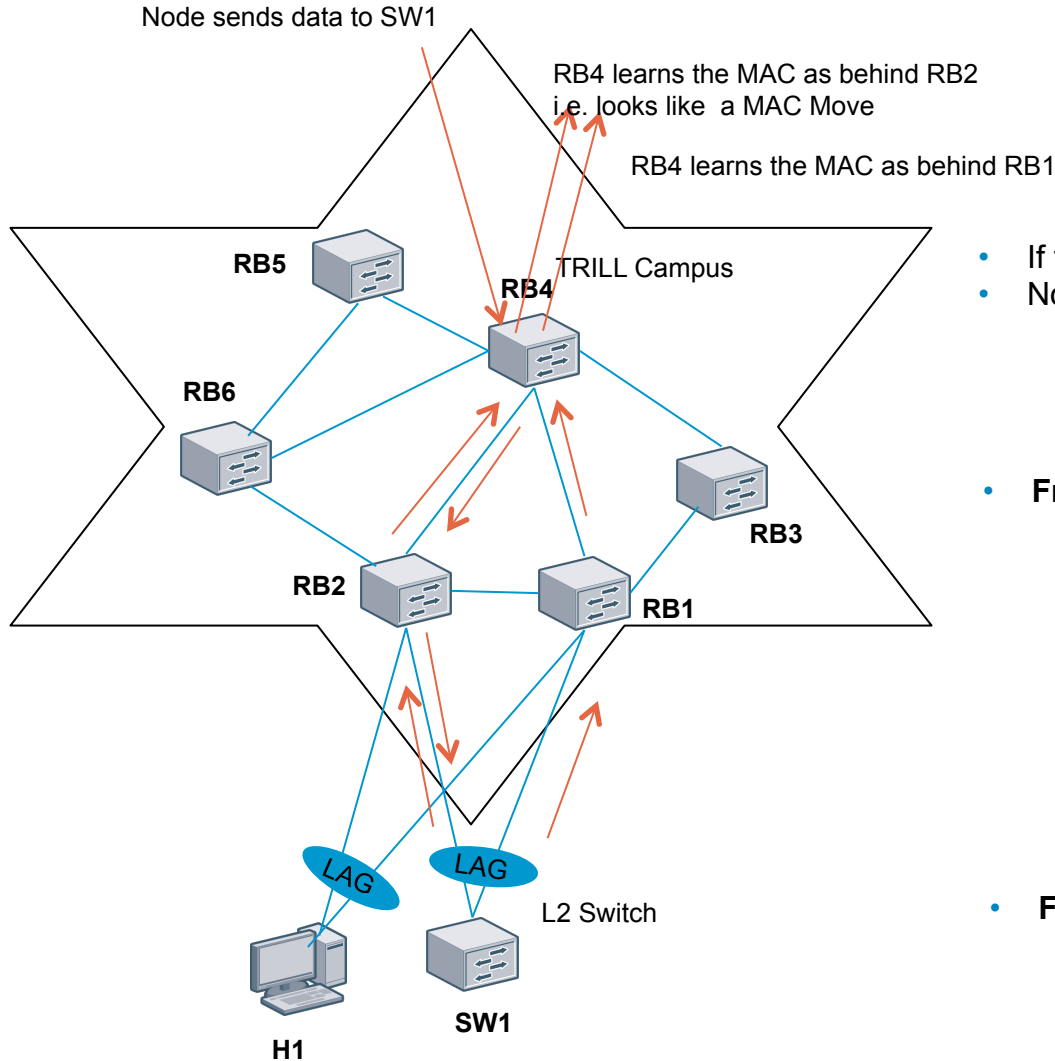
---

# Handling of Multihoming in TRILL



- **Switch Multi-homing to TRILL Campus (SW1)**
  - RB1, RB2 will elect a DRB and AF for VLANs on that link
  - RB1 gets selected as the AF for the link
  - RB1 – SW1 link will be used for traffic in and out of the TRILL Campus and SW1 on that set of VLANs
- **End host multi-homing to TRILL campus**
  - RB1 and RB2 will not see each other through the host
  - In this configuration, the host should operate in Active-Passive mode, i.e. it should send and accept the data from only one link
  - This link can be RB1 – H1 link
  - If this link fails, then H1 will switch to the other link, which is H1-RB2 link

# Handling of Multihoming in TRILL

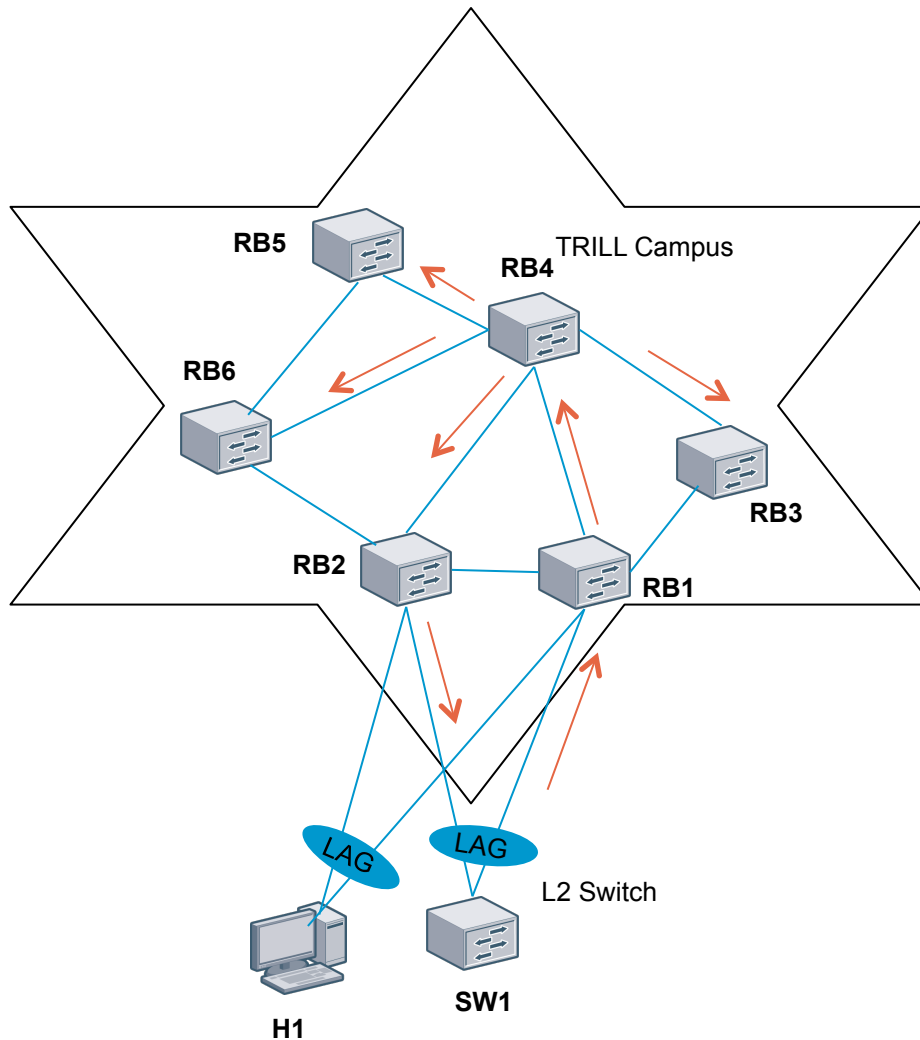


- If the external node treats the **two links as a LAG**
- No AF formed between RB1 and RB2

- **Unicast Data Handling**

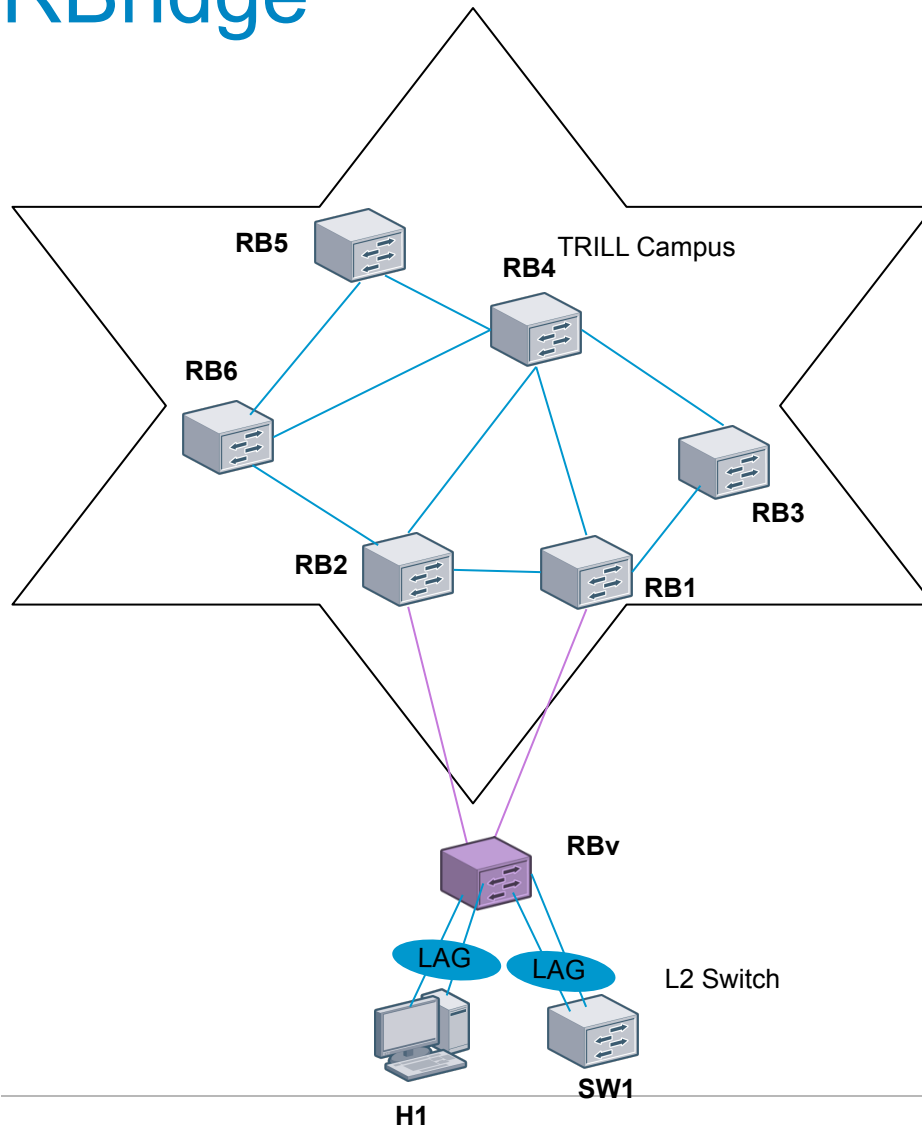
- **From SW1 to the network**
  - If packet is sent through RB1, RB4 learns the MAC as belonging to RB1
  - Later when due to hashing another flow is sent through RB2, RB4 sees this as a MAC move
  - Results in constant MAC Move and also potential packet re-ordering
  - Can be avoided if SW1 uses MAC based or VLAN based hash distribution
- **From network to SW1**
  - If RB2 is the last seen RBridge for this MAC, packet is sent to RB2
  - No Multipathing between RB1 and RB2 from the network

# Handling of Multihoming in TRILL



- If the external node treats the **two links as a LAG**
- No AF formed between RB1 and RB2
- **Multicast Data Handling**
- **Remove AF for Multicast Traffic**
  - SW1 will receive duplicate traffic
  - Traffic sent from SW1 to RB1, will again be reflected back to it on the RB2-SW1 link
  - Traffic from the network, would get delivered twice, once through RB1 and next time through RB2

# Handling of Multihoming in TRILL – Virtual RBridge

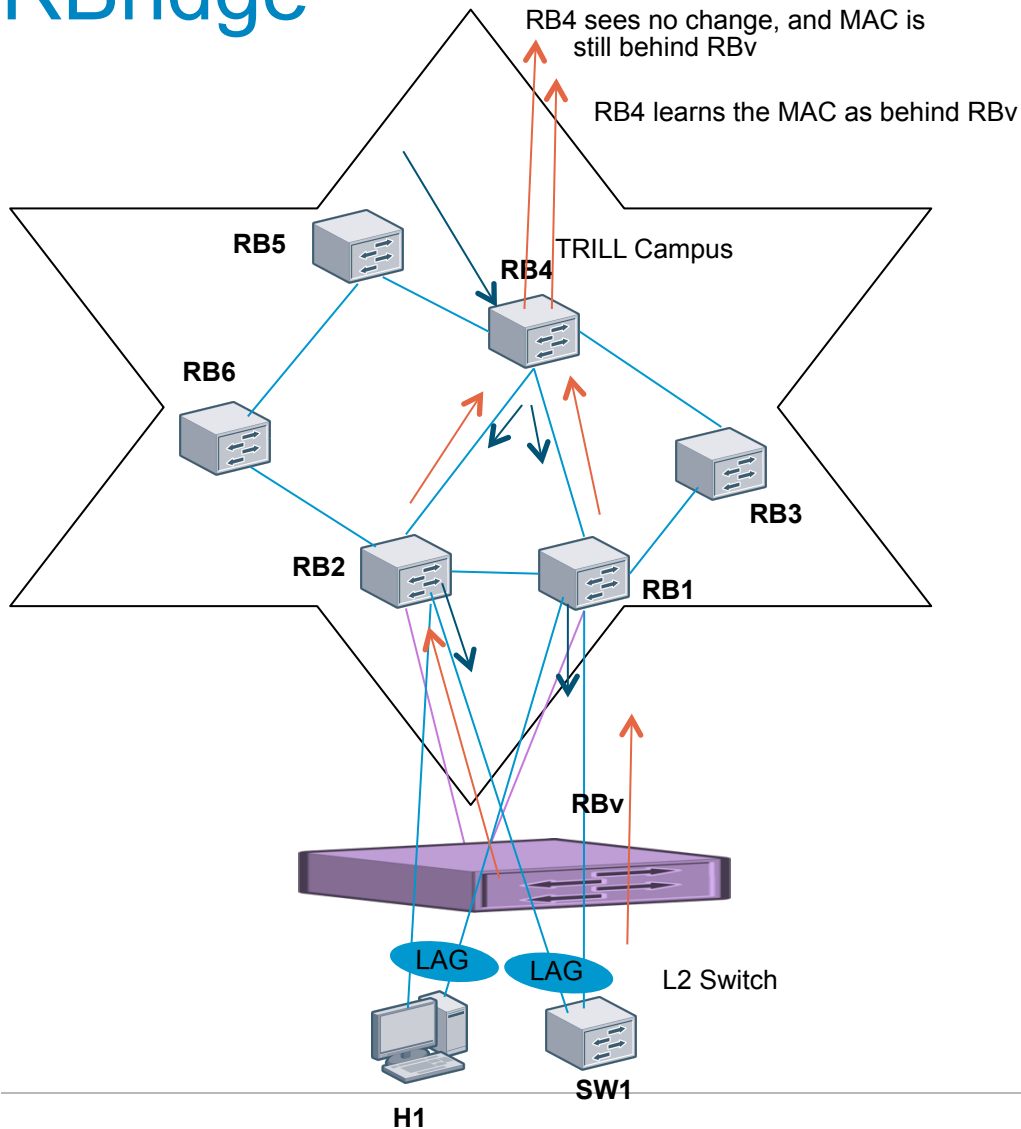


- Virtual RBridge
  - The physical topology remains the same as before
  - The RBridges RB1 and RB2 generate LSP as though there is a RBridge RBv, to which both of them are connected
  - This is called the virtual RBridge
  - Logically the network diagram looks as shown with the virtual RBridge, and the external host and switch connected to this Virtual RBridge
  - There can be multiple nodes behind the same Virtual RBridge
  - Useful for scaling compared to having one Virtual RBridge per connected node
- Behavior of the external nodes
  - The external nodes need to treat their links to the network as a LAG
  - If using Dynamic LAG (LACP) will need coordination between RB1 and RB2 (This is out of scope of this proposal)

# Handling of Multihoming in TRILL – Virtual RBridge

- ~~Realization of Virtual RBridge~~ traffic between RB1 and RB2 through RBv
  - › In LSP for RBv, set the IS-IS overload bit on
    - Ensure RBv cannot be a root RBridge for multicast trees
  - › Creation of the Virtual RBridge
  - ›
    - This will be because RBridge is multiple nodes which are named
  - › Could be by configuration or some dynamic learning protocol
- Out of scope of this proposal
  - › Both RB1 and RB2 must agree on the Virtual RBridge ID to be used for this purpose
    - Could be by configuration or some dynamic protocol between the two
  - Out of scope of the proposal
    - Out of scope of this proposal
  - › Both RB1 and RB2 must agree on the Virtual RBridge ID to be used for this purpose
    - Could be by configuration or some dynamic protocol between the two
    - Out of scope of the proposal

# RBridge



## Handling of Unicast data traffic

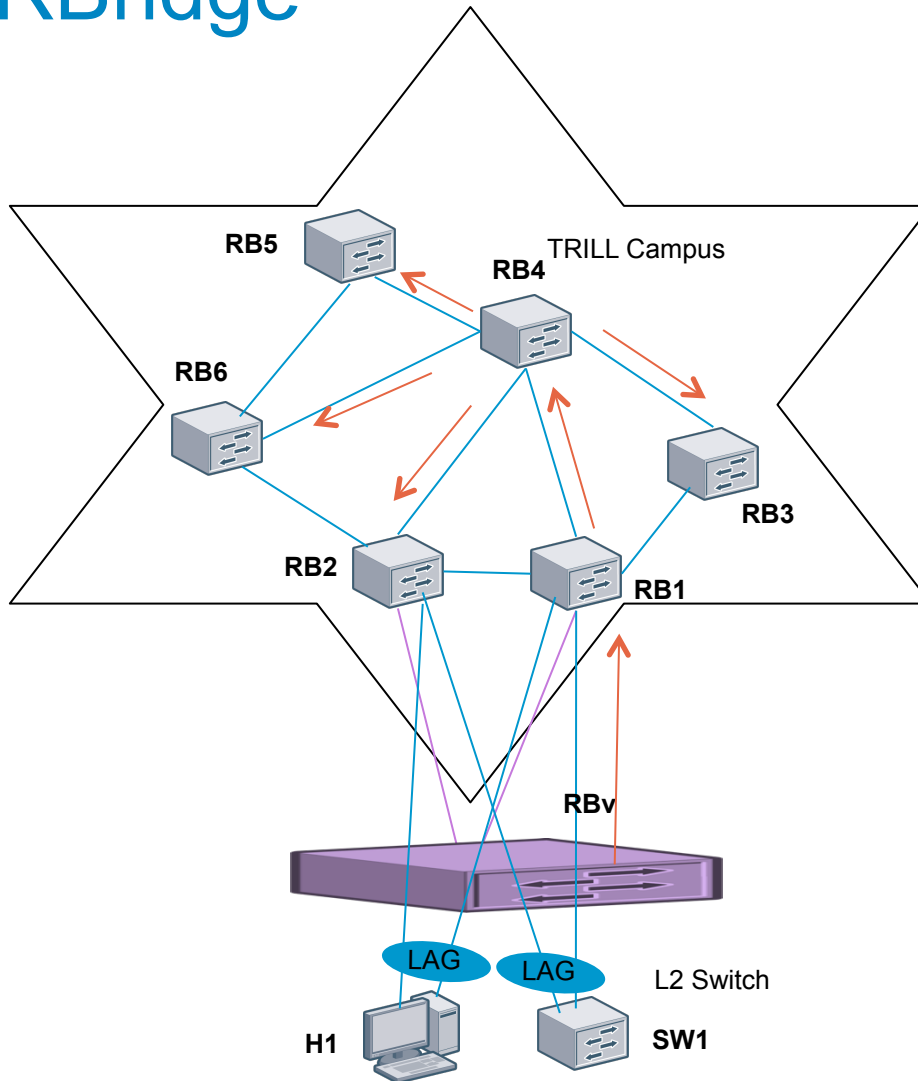
### Data from External host/Switch to the network

- SW1 sends traffic to RB1
- RB1 knows that these ports are used for dual homing, and the associated Virtual RBridge is RBv
- Sends the TRILL frame, with ingress RBridge set to RBv
- RB4 now learns the MAC as being behind RBv
- SW1 due to hashing sends packet to RB2
- RB2 also knows that these ports are used for dual homing, and the associated Virtual RBridge is RBv
- Sends the TRILL frame with ingress RBridge set to RBv
- RB4 now does not see this as MAC move, and no problems

### Data from Network to the external Switch/Host

- RB4 gets data for SW1, and the lookup shows the MAC as reachable through RBv
- RB2
- Uses ECMP and distributes traffic via RB1 and RB2
- When RB1/RB2 receive traffic, they realize that RBv is owned by them (Virtual RBridge they created)
- Decapsulate and forward the frame to the port

# Handling of Multihoming in TRILL – Virtual RBridge



## Handling Flooding data traffic

- Mandate use of minimum of two trees from each root node
- One of the trees should be engineered to choose RB2 – RBv link, the other tree should be engineered to choose RB1 – RBv link
- This is uniform for all Rbridges in the network
- In the example, it is RB4-1 (active link RB1-RBv), and RB4-2 (active link is RB2-RBv)
- If RB1 is a root bridge, then form another tree with RB2 as the root bridge

## Data from External host/Switch to the network

- SW1 uses some load balancing and forwards the traffic to RB1
- RB1 will send it on TRILL N/W Ingress RBridge as RBv, and Tree as RB4-1 (or any tree where RBv-RB1 link is active)
- When traffic reaches RB2, RB2 will not forward to ports which belong to VRb, since it will be pruned for the tree RB4-1

## Data from External host/Switch to the network

- SW1 uses some load balancing and forwards the traffic to RB2
- RB2 will send it on TRILL N/W Ingress RBridge as RBv, and Tree as RB4-2 (or any tree where RBv-RB2 link is active)

- When traffic reaches RB1, RB1 will not forward to

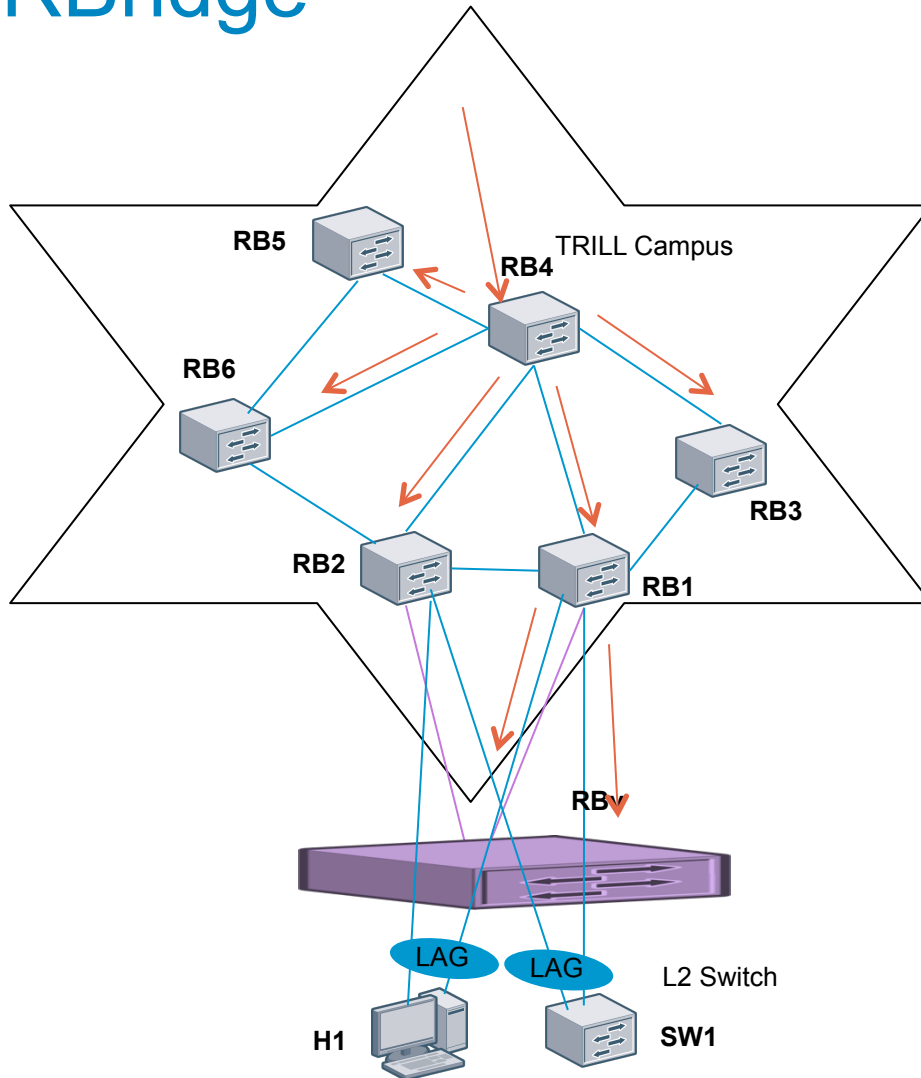
## Data from External host/Switch to the network

- for the tree RB4-1



# Handling of Multihoming in TRILL – Virtual RBridge

## Handling Flooding data traffic



### Data from Network

- RB4, which is the ingress RBridge chooses either RB4-1 or RB4-2 as the tree
- If it chooses RB4-1, as the distribution tree, then RB1 will forward the packets out to the ports belonging to RBv
- If RB4 alternatively chooses RB4-2 as the distribution tree, then RB2 will forward it to the ports of RBv, and RB1 will drop it on these ports

# RBridge

## – Requirements from the Hardware/Data Plane

### › Nickname

- › Ability to support multiple nickname

›

Ability to set the ingress RBridge to one of the nicknames associated with this

- › Ability to set the ingress RBridge to one of the nicknames associated with this RBridge receiving data, should be able to recognize multiple nicknames as being associated with this RBridge

access ports (VLAN flooding) Virtual RBridge ports is associated with that Virtual RBridge and belonging to that VLAN

## – Multicast Data Forwarding

›

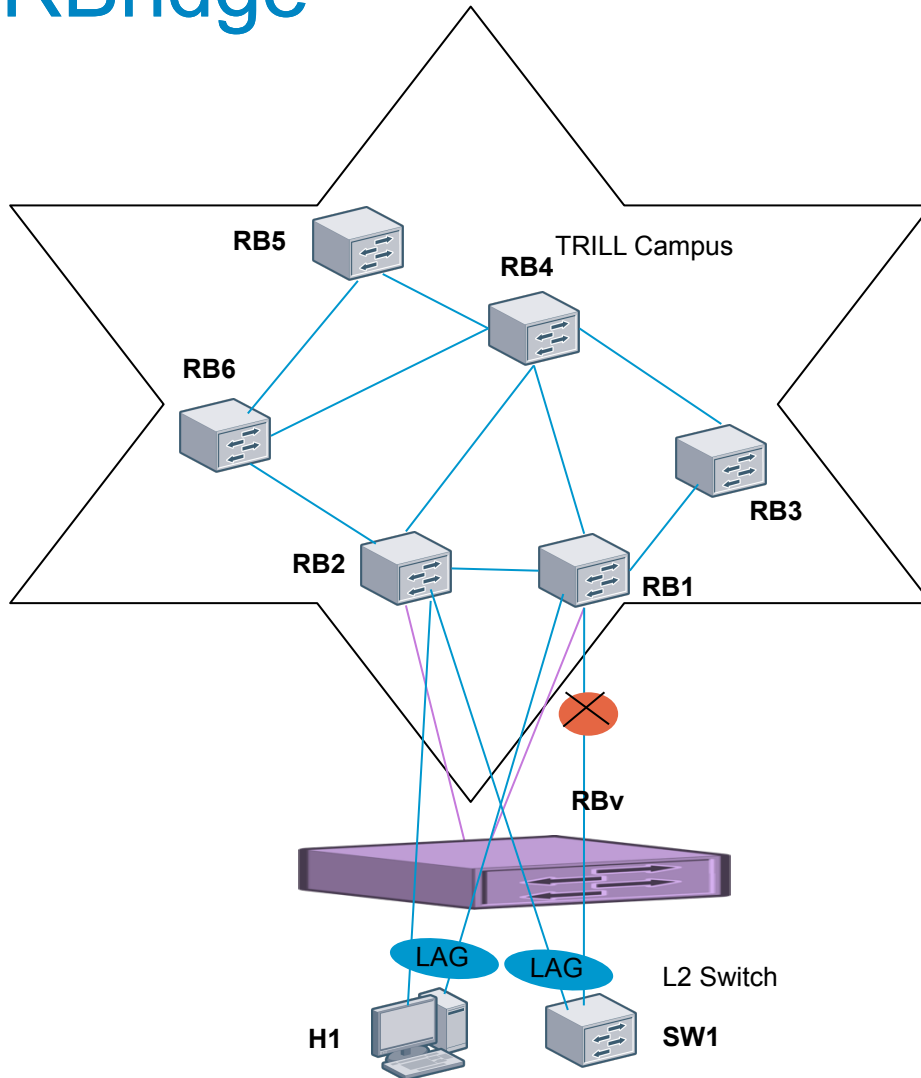
For a tree, for which the link between this RBridge and the virtual RBridge is not

active, prune all the ports of the virtual RBridge

a per port basis for the link between this RBridge and the virtual RBridge is not active, prune all the ports of the virtual RBridge

- › Ability to choose the ingress RBridge and the distribution tree to use at least on a per port basis

# RBridge



Link between SW1 and RB1 fails

- RB1 informs RB2 of this failure
- RB2 knows the corresponding Link between RB2 and SW1
- RB2 will move that link out of the set of ports controlled by VRb
- Further packets from this port, will be encapsulated with ingress RBridge set as RB2
- The rest of the network, sees this as a MAC move
- From now on, the packets will be sent to RB2

Link becomes Operational again

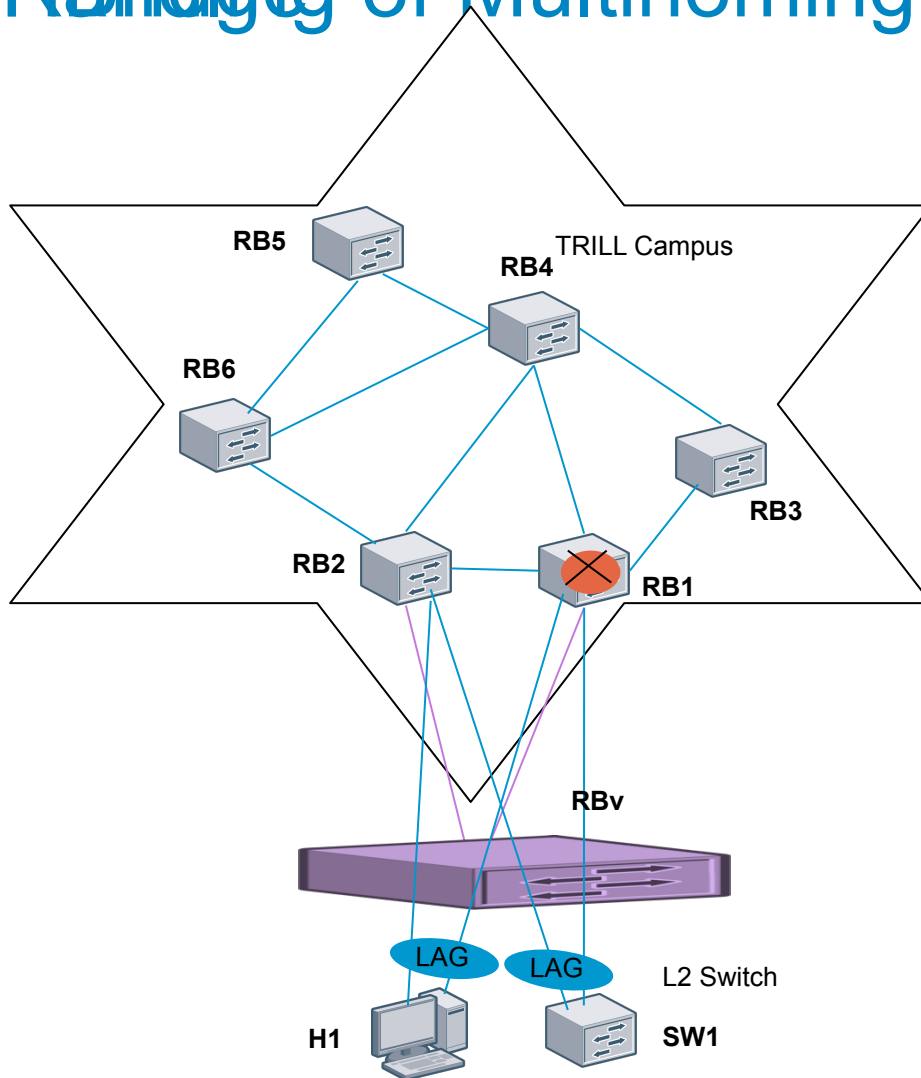
- RB1 should inform RB2 of the link up
- RB1 will then revert to using VRb as the ingress RBridge for this port

# Handling of Multihoming in TRILL – Virtual Bridging

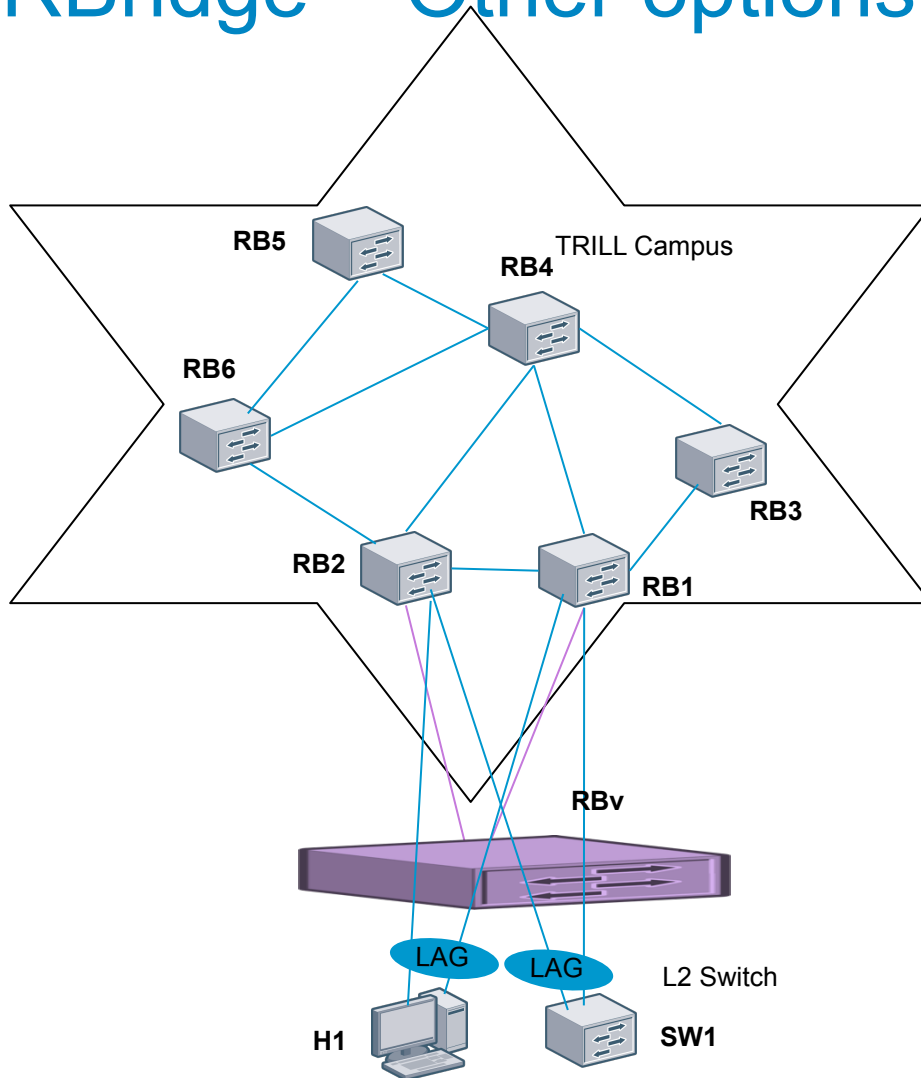
## Handling Node Failure

RB1 fails

- If RB1 is the flooding packet forwarder for VRb, then RB2 should take over this role
- RB2 will continue to advertise these ports using the VRb, so that when RB1 comes back, there is no MAC move operations



# Handling of Multihoming in TRILL – Virtual RBridge – Other options for multicast data



## Handling Flooding data traffic

- In every tree of the network, either RB2 – RBv will be active or RB1-RBv will be active
- One option is to insist on trees with RB1 and RB2 as root, but will not scale for if there are lots of RBv

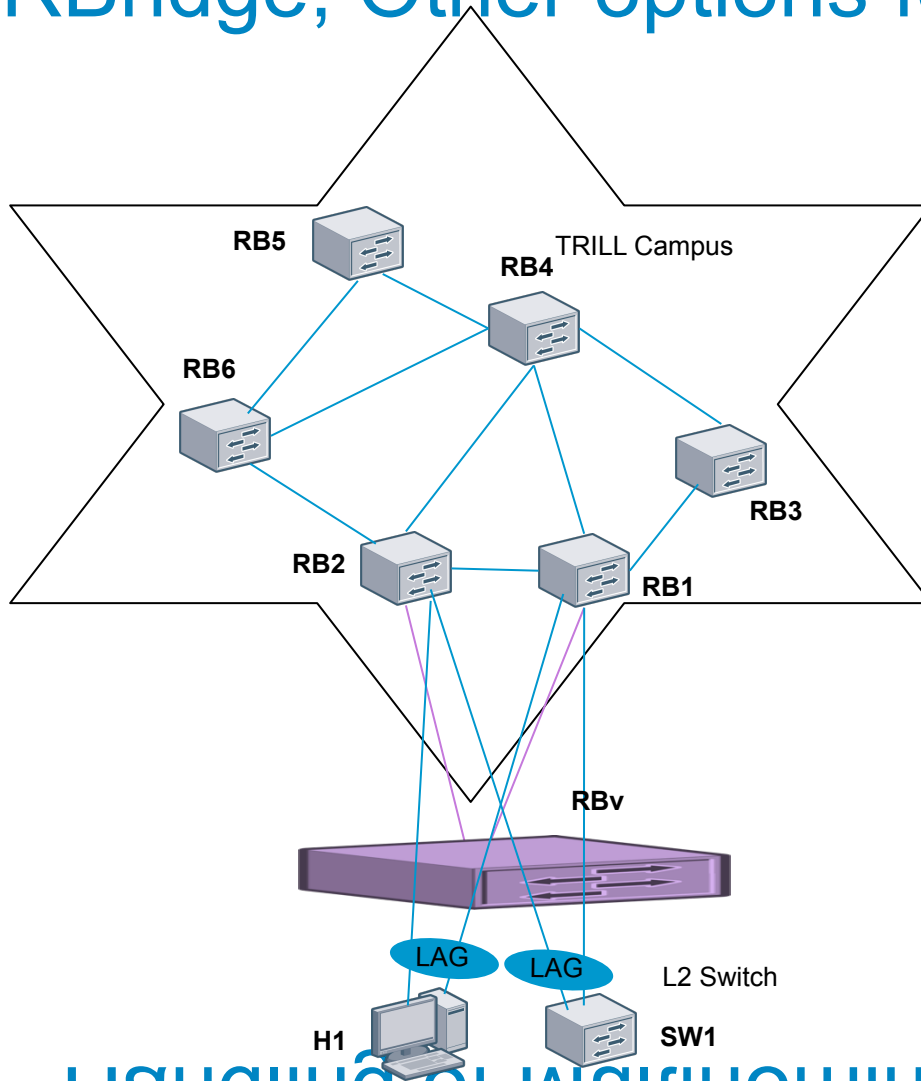
## Handling Traffic from Network

- The ingress node would have picked a distribution tree
- In this distribution tree, either RB1 or RB2 would have link to RBv
- The other RBridge, should prune the ports of RBv from this distribution tree
- Hence RBv ports will get this from packet only once either from RB1 or RB2

## Handling Traffic from SW1 to network

- The packet can reach either RB1 or RB2
- If there is a tree with this RB – RBv as active link, use that as the distribution tree for these ports
- If there is no such tree, then tunnel the packet to the other RB and have that forward the packet to the network ???

# Handling of Multihoming in TRILL – Virtual RBridge, Other options for multicast data



## Handling Flooding data traffic

- In every tree of the network, either RB2 – RBv will be active or RB1-RBv will be active

## Handling Traffic from Network

- The ingress node would have picked a distribution tree
- In this distribution tree, either RB1 or RB2 would have link to RBv
- The other RBridge, should prune the ports of RBv from this distribution tree
- Hence RBv ports will get this from packet only once either from RB1 or RB2

## Handling Traffic from SW1 to network

- The packet can reach either RB1 or RB2
- If there is a tree with this RB – RBv as active link, use that as the distribution tree for these ports
- Send the packet with Ingress RBridge as RBv, but need to relax the RPF check  
Could be costly in case of temporary loops