Tunnel Stitching for Overlay Traffic Transport

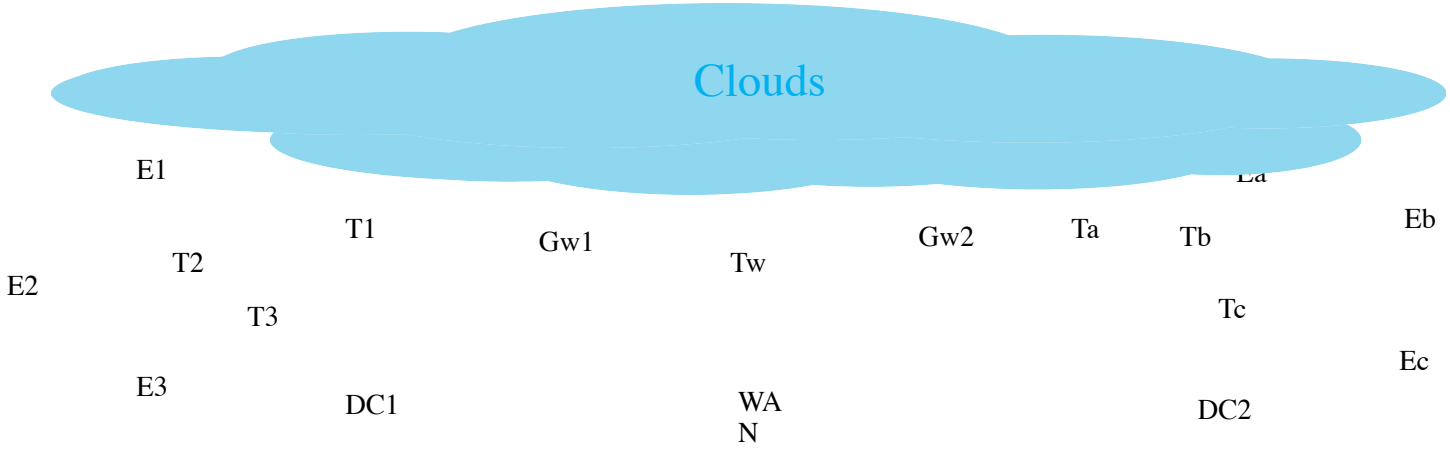
draft-yong-nvo3-tunnel-stitching

Lucy Yong lucy.yong@huawei.com
Weiguo Hao haoweiguo@huawei.com

July 2016,
Berlin

About Cloud Traffic Transport

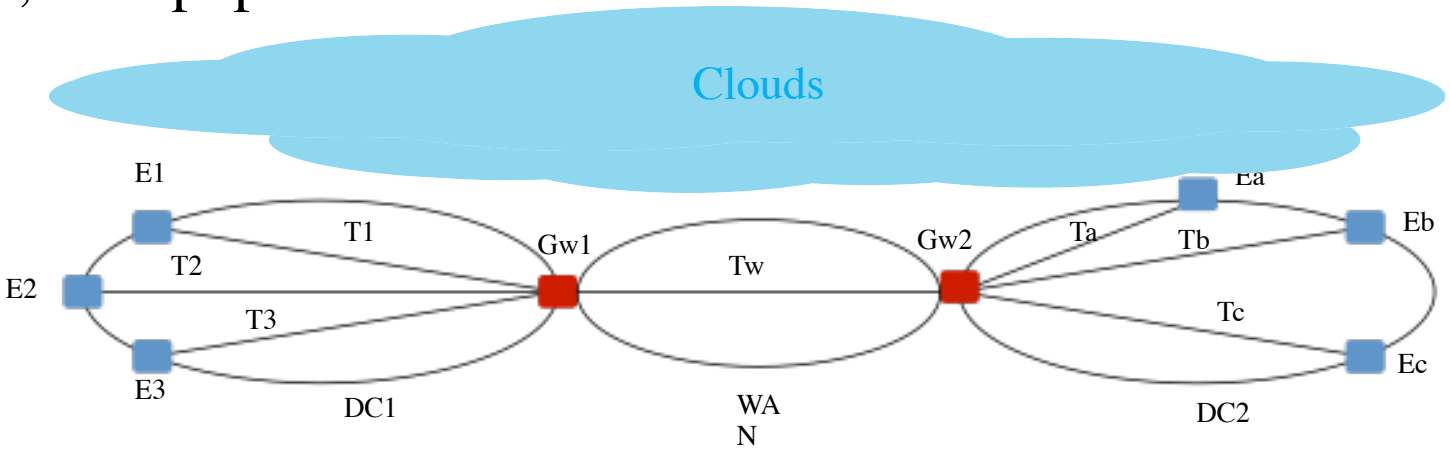
- Cloud traffic may traverse multiple network segments
- Cloud traffic is often known as overlay traffic
- Overlay traffic is often delivered via an IP tunnel
- Overlay traffic over multiple networks, each of them use an IP tunnel to delivery overlay traffic



Tunnel Stitching Example

Overlay Traffic Delivery

- E1, E2, E3 encapsulate incoming traffic and send encapsulated packets to Gw1 over T1,T2,T3
- Gw1 receives IP packets from E1, E2, E2, decapsulates the pacekts, performs payload dst address lookup to find next tunnel end point, i.e. GW2; encap payload again, forward
- Gw2 receives IP packets from Gw1, decaps., perform payload dst address lookup to find next tunnel end point, i.e. Ta, Tb, or Tc; Encap. packets and forward

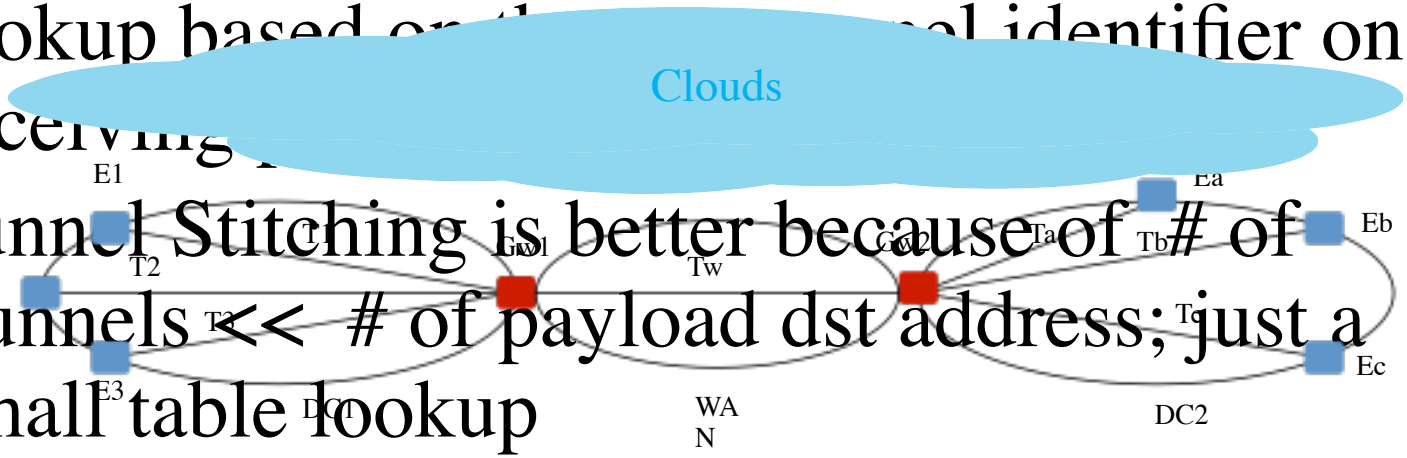


Issue for Overlay Traffic Delivery

- When Overlay traffic traverse multiple network segments,
  - A board node has to terminate previous tunnel, then performs payload lookup, and constructs a next tunnel
  - Operation at a board node is complex, not scale, increases overlay traffic e2e delay
    - Cloud address can be very scatter, hard to aggregate -> lookup table will be huge -> lookup time will be long
    - One board node may serve for many cloud instances -> each has a big table

Tunnel Stitching

- Tunnel Stitching is the technique to pass overlay traffic from one tunnel to next tunnel without a payload dst lookup
- Tunnel Stitching is to encode the next tunnel identifier into a previous tunnel encap. header
- A board node performs next tunnel endpoint lookup based on the tunnel identifier on receiving
- Tunnel Stitching is better because of # of Tunnels << # of payload dst address; just a small table lookup
  - Reduce board node burden , shorten overlay



Solution Works

- Enhancement in NVO3 encapsulation header to encode next tunnel identifier
  - VXLAN, VXLAN-GRE, GUE, Geneve, etc
- Enhancement on NVA for SDN implementation
  - NVA pushes <overlay dst address, tunnel IP address, next tunnel identifier> to first NVE
  - NVA pushes <next tunnel identifier, tunnel IP address, next-next tunnel identifier> to the stitching nodes
- Enhancement on BGP for distributed implementation