# NVO3 Requirements for Tunneling

Igor Gashinsky and Bruce Davie
IETF

# Why tunnels?

- Manage overlapping addresses between multiple tenants
- Decouple virtual **topology** provided by tunnels from physical network topology
- Decouple virtual **network service** from physical network (e.g. provide an L2 service over an L3 fabric)
- Support VM mobility independent of the physical network
- Support larger numbers of virtual networks (vs. VLANs for example)
- Reduce state requirements for physical network (e.g. MAC addresses)
- Because all CS problems can be solved with another level of indirection

# Disclaimers

- We have a horse in this race, but we've tried hard to be objective
- AFAICT, no existing protocol meets all the requirements in this presentation
- We've put a lot of emphasis on compatibility with existing HW – others may differ on the importance of that

# Requirements Overview

- Control Plane independence
- Backwards compatibility
  - Lots of installed devices & services to consider
- Context identification

# Control Plane Independence

- Data planes tend to get baked into HW, control planes evolve
  - Best not to specify control plane as part of tunnel encaps

# Backwards Compatibility (1)

- With switches and routers
  - IP-based encaps likely to be most compatible
  - ECMP – mostly looks at IP src/dst and TCP or UDP ports, so make use of that
- With NICs
  - Most tunneling methods break TSO, causing major performance hit for host-terminated tunnels
  - For current generation NICs, only way to keep TSO is to completely match TCP header – see draft-davie-stt-01

# Backwards Compatibility (2)

- Middle Boxes
  - Should be possible to transit them
  - They may need to inspect payload (e.g. for stateful firewall)
  - Stream or Frame Reassembly may be needed for L4/L7 services
- Hardware or software-based “NV Edge”
  - “Edge” may be in hypervisor, physical switch, appliance etc.
- With WAN services (e.g. Public IP, L3VPNs, VPLS)
  - These services carry IP or Ethernet, so compatible with IP-based encaps

# Context Identification

- As packets exit from tunnels, need to deliver them to the right "context"
  - A context may be simply a "tenant", or a "virtual network instance" but these are special cases
  - Can also use it for other metadata (state versioning, distributed lookup, etc.)
  - Note that L3VPNs don't have any single field that is the VPN-ID, and that's a good thing
    - Allows much more complex notions of VPN membership than "a member of exactly one VPN"
  - An opaque context ID with control-plane defined semantics also supports control-plane independence goal