Network Virtualization Overlay Control Protocol Requirements

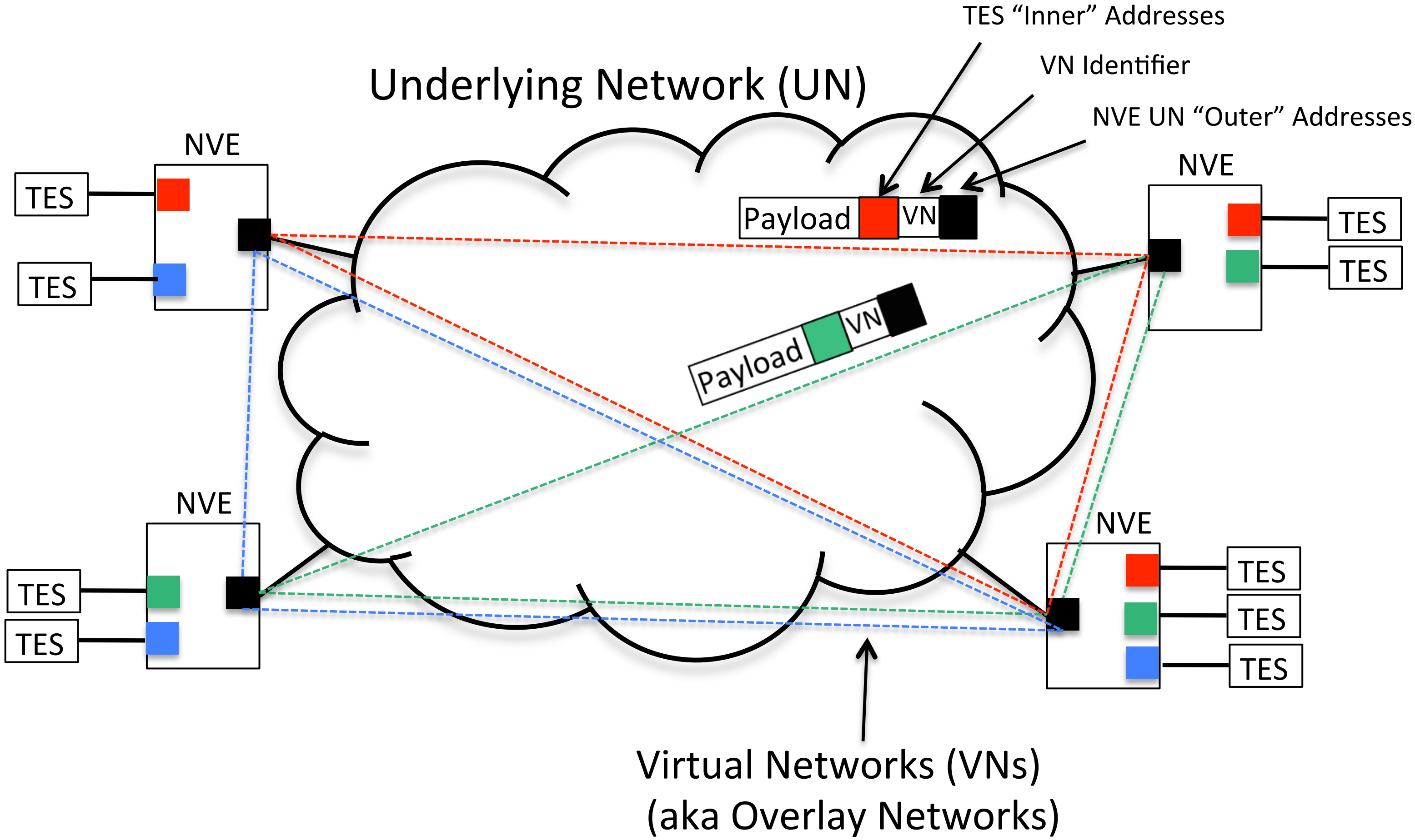
draft-kreeger-nvo3-overlay-cp-00

Lawrence Kreeger, Dinesh Dutt, Thomas Narten, David Black, Murari Sridharan

Purpose

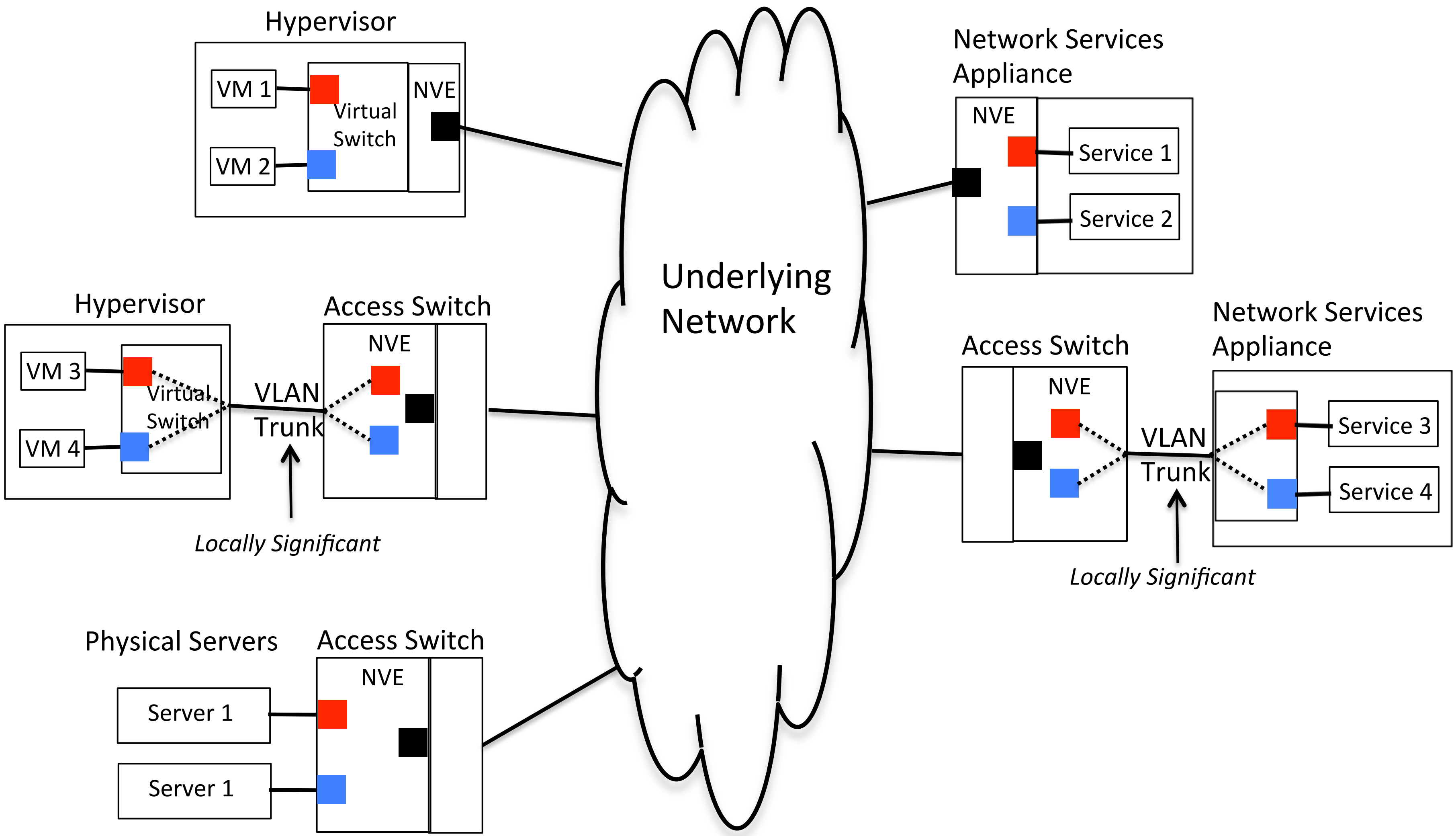
Outline the high level requirements for control protocols needed for overlay virtual networks in highly virtualized data centers.

Basic Reference Diagram



Network Virtualization Edge (NVE) – (OBP in draft)
Tenant End System (TES) – (End Station in draft)

Possible NVE / TES Scenarios



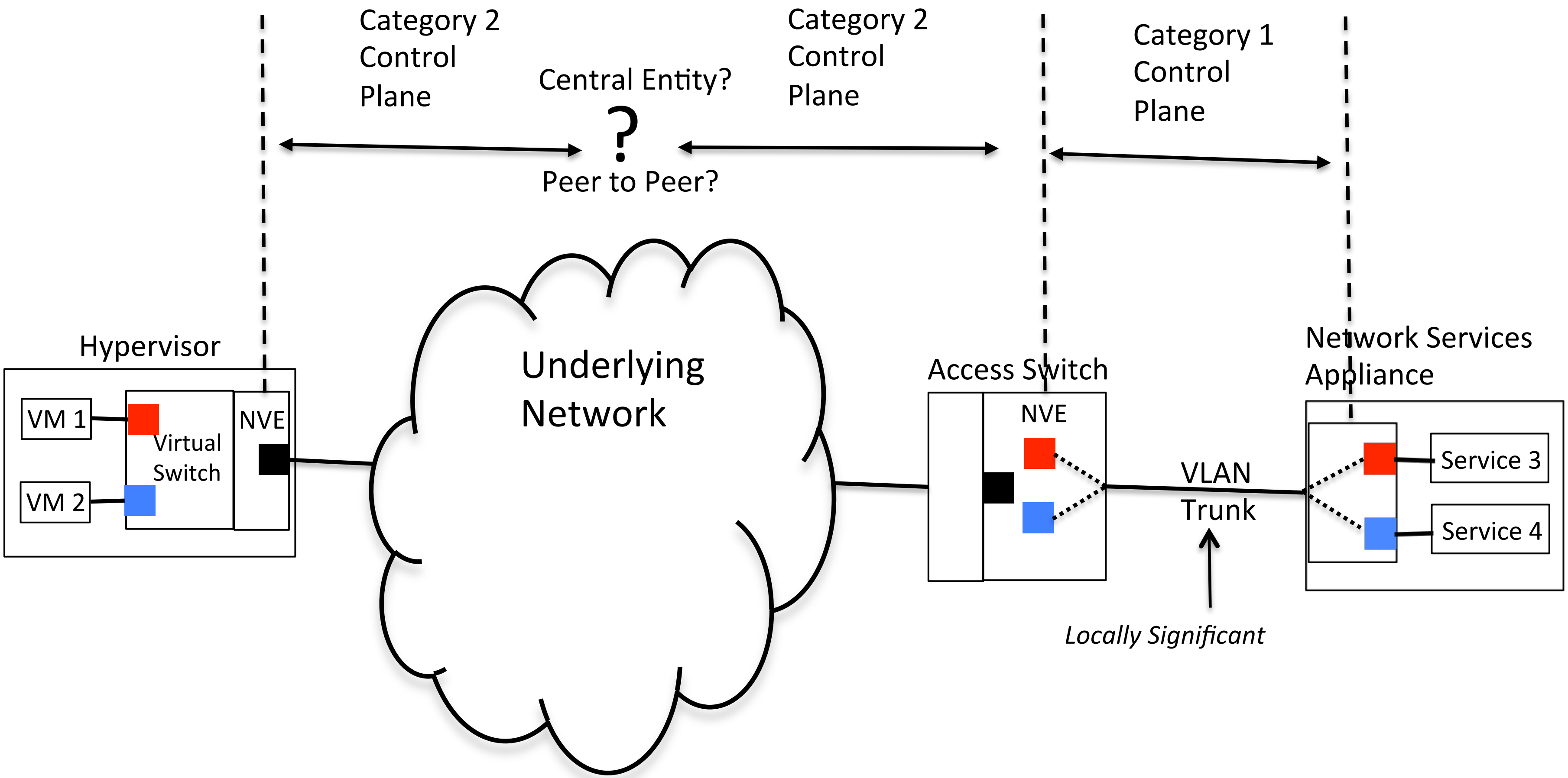
Dynamic State Information Needed by an NVE

- Tenant End System (TES) inner address (scoped by Virtual Network (VN)) to outer (Underlying Network (UN)) address of the other Network Virtualization Edge (NVE) used to reach the TES inner address.
- For each VN active on an NVE, a list of UN multicast addresses and/or unicast addresses used to send VN broadcast/multicast packets to other NVEs forwarding to TES for the VN.
- For a given VN, the Virtual Network ID (VN-ID) to use in packets sent across the UN.
- If the TES is not within the same device as the NVE, the NVE needs to know the physical port to reach a given inner address.
  - If multiple VNs are reachable over the same physical port, some kind of tag (e.g. VLAN tag) is needed to keep the VN traffic separated over the wire.

Two Main Categories of Control Planes

1. For an NVE to obtain dynamic state for communicating with a TES located on a different physical device (e.g. hypervisor or Network Services Appliance).
2. For an NVE to obtain dynamic state for communicating across the Underlying Network to other NVEs.

Control Plane Category Reference Diagram

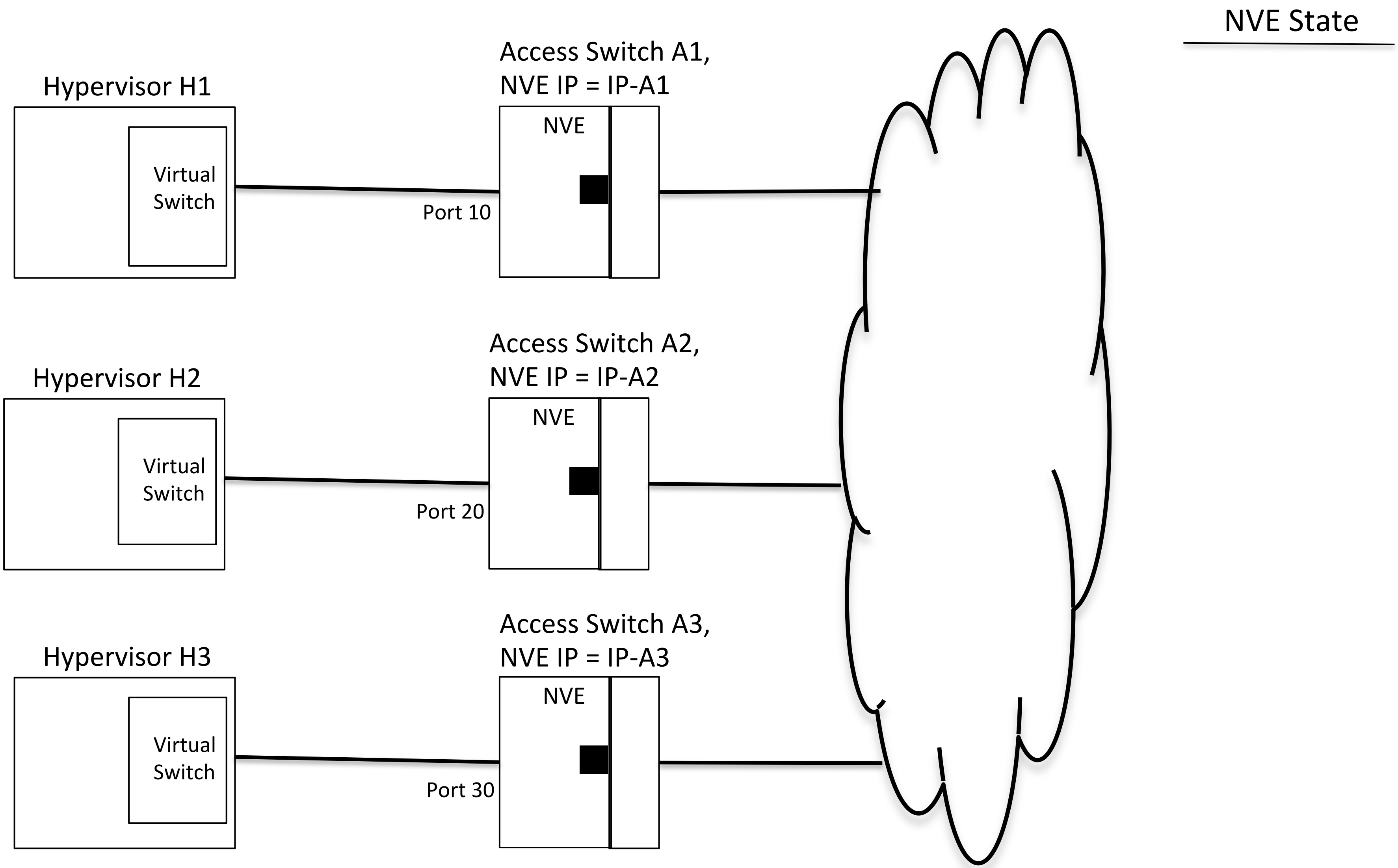


Category 2 CP Architecture Possibilities

- Central entity is populated by DC orchestration system
- Central entity is populated by Push from NVE
- Push to NVE from central entity
- Pull from NVE from central entity
- Peer to Peer exchange between NVEs with no central entity
- Central entity could be a monolithic system or a distributed system

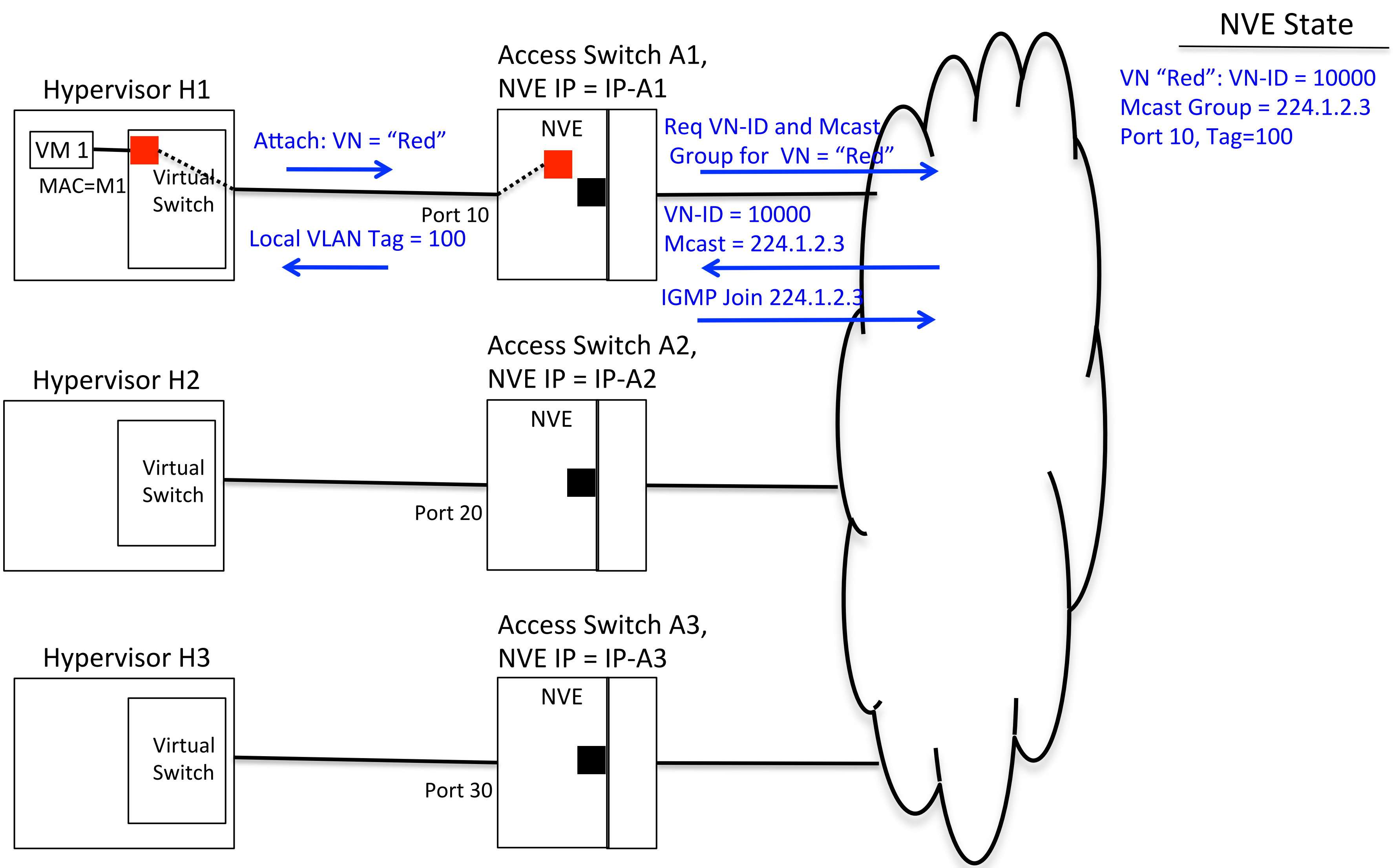
Possible Example CP Scenario

This example is not part of the Req draft and is shown for illustrative purposes
Assumes: Central entity with push/pull from NVE, Multicast Enabled IP Underlay



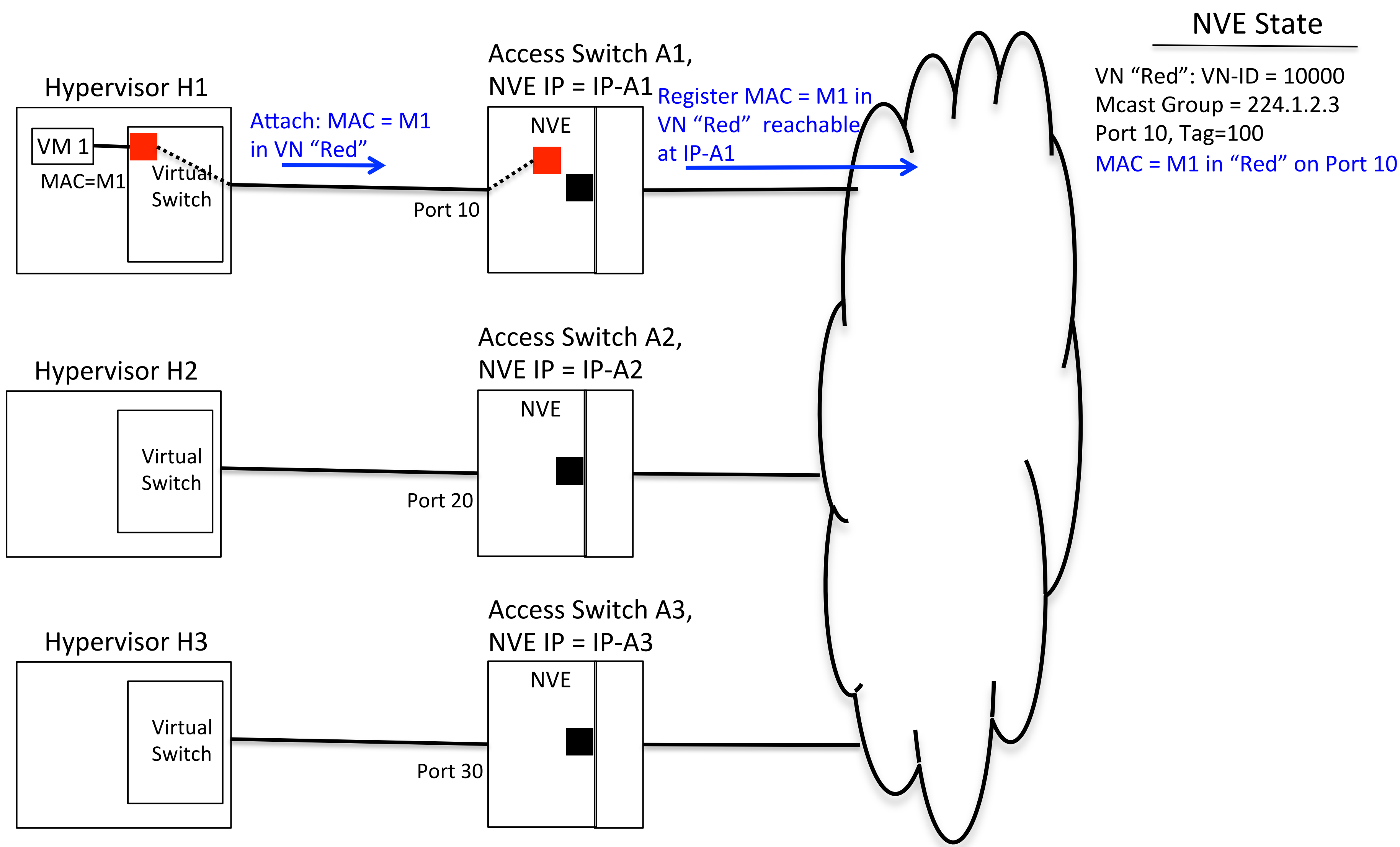
VM 1 comes up on Hypervisor H1, connected the VN “Red”

H1’s Virtual Switch signals to A1 that it needs attachment to VN “Red”



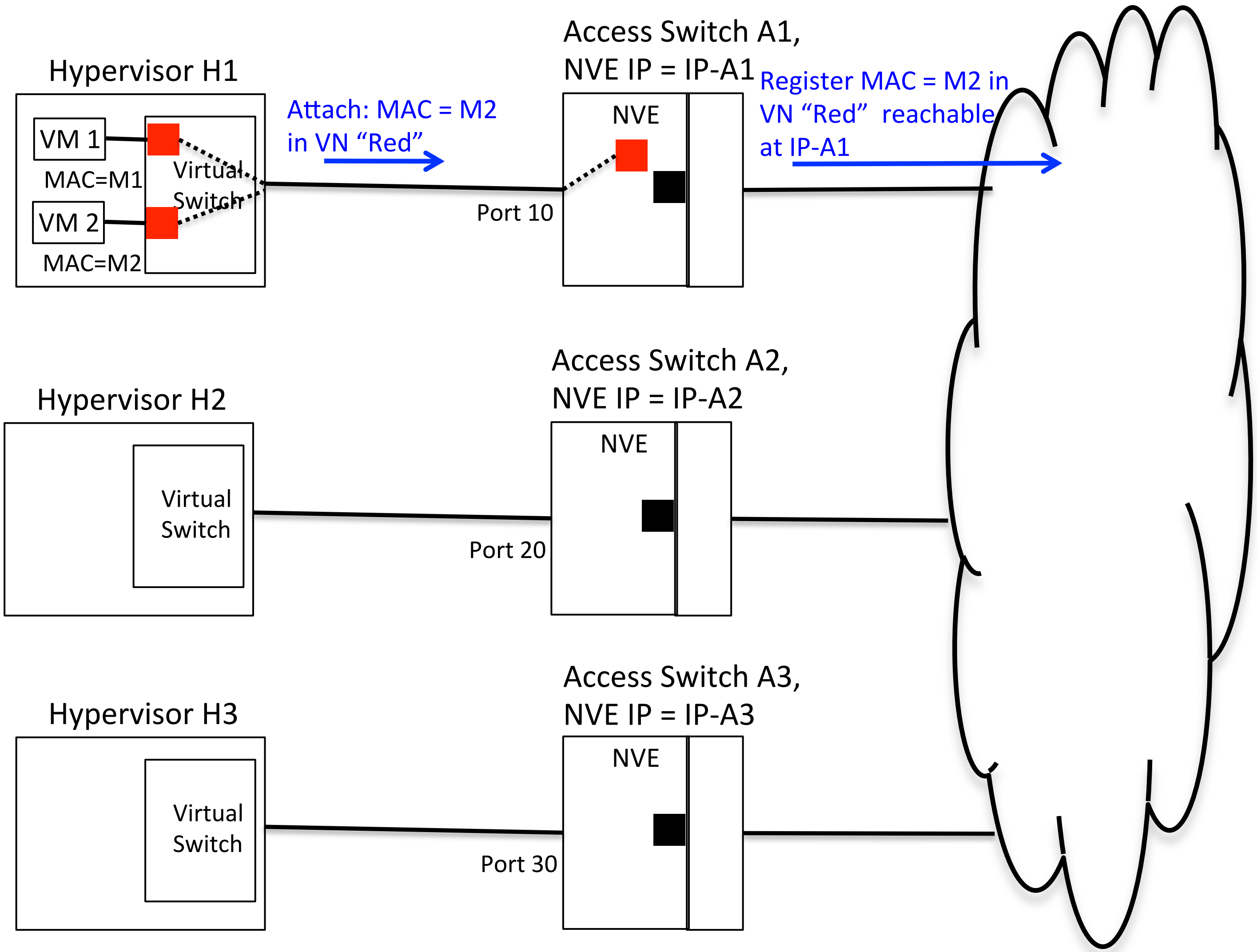
VM 1 comes up on Hypervisor H1, connected the VN “Red”

H1’s Virtual Switch signals to A1 that MAC M1 is connected to VN “Red”



VM 2 comes up on Hypervisor H1, connected the VN “Red”

H1’s Virtual Switch signals to A1 that MAC M2 is connected to VN “Red”

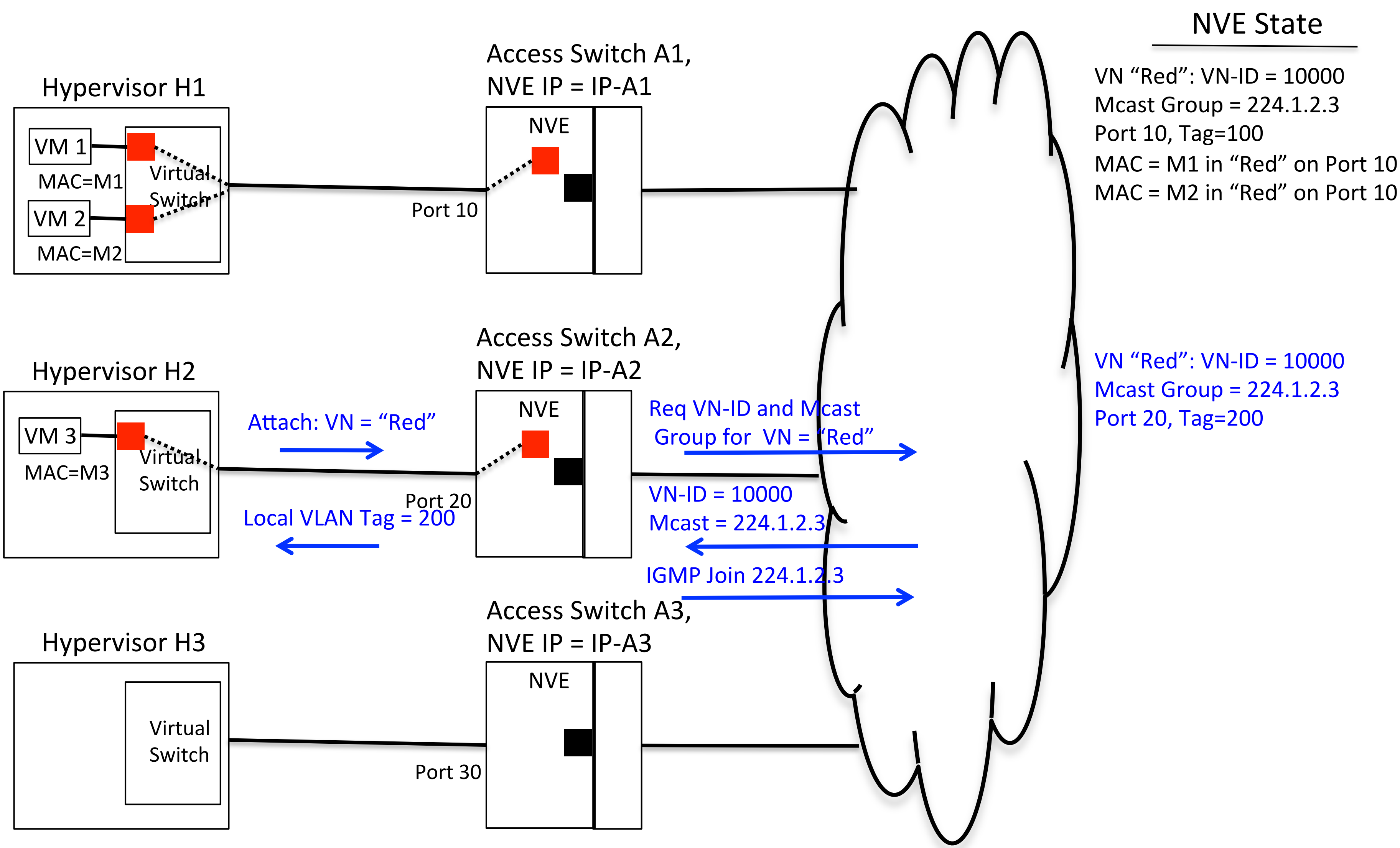


NVE State

VN “Red”: VN-ID = 10000
Mcast Group = 224.1.2.3
Port 10, Tag=100
MAC = M1 in “Red” on Port 10
MAC = M2 in “Red” on Port 10

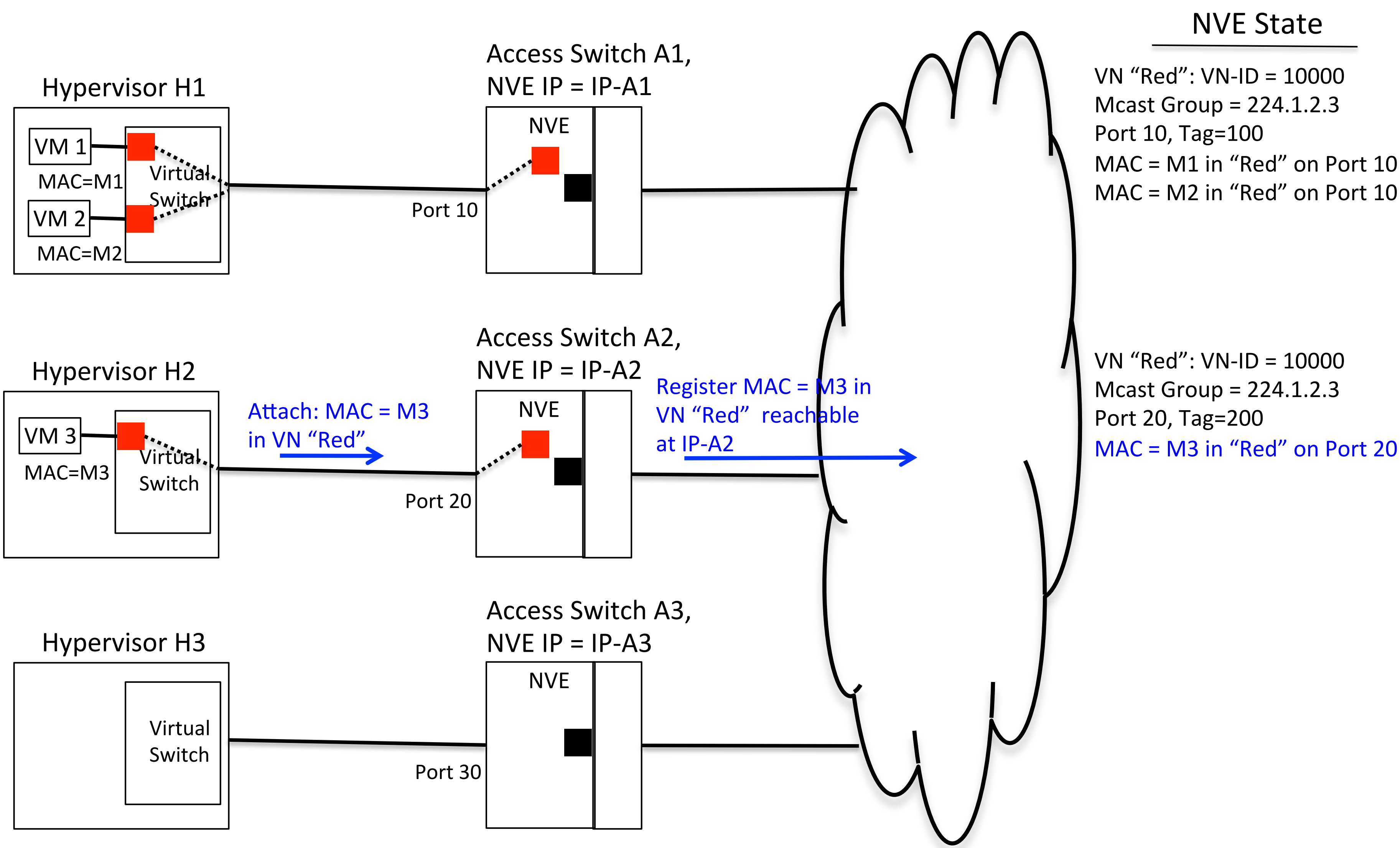
VM 3 comes up on Hypervisor H2, connected the VN “Red”

H2’s Virtual Switch signals to A2 that it needs attachment to VN “Red”



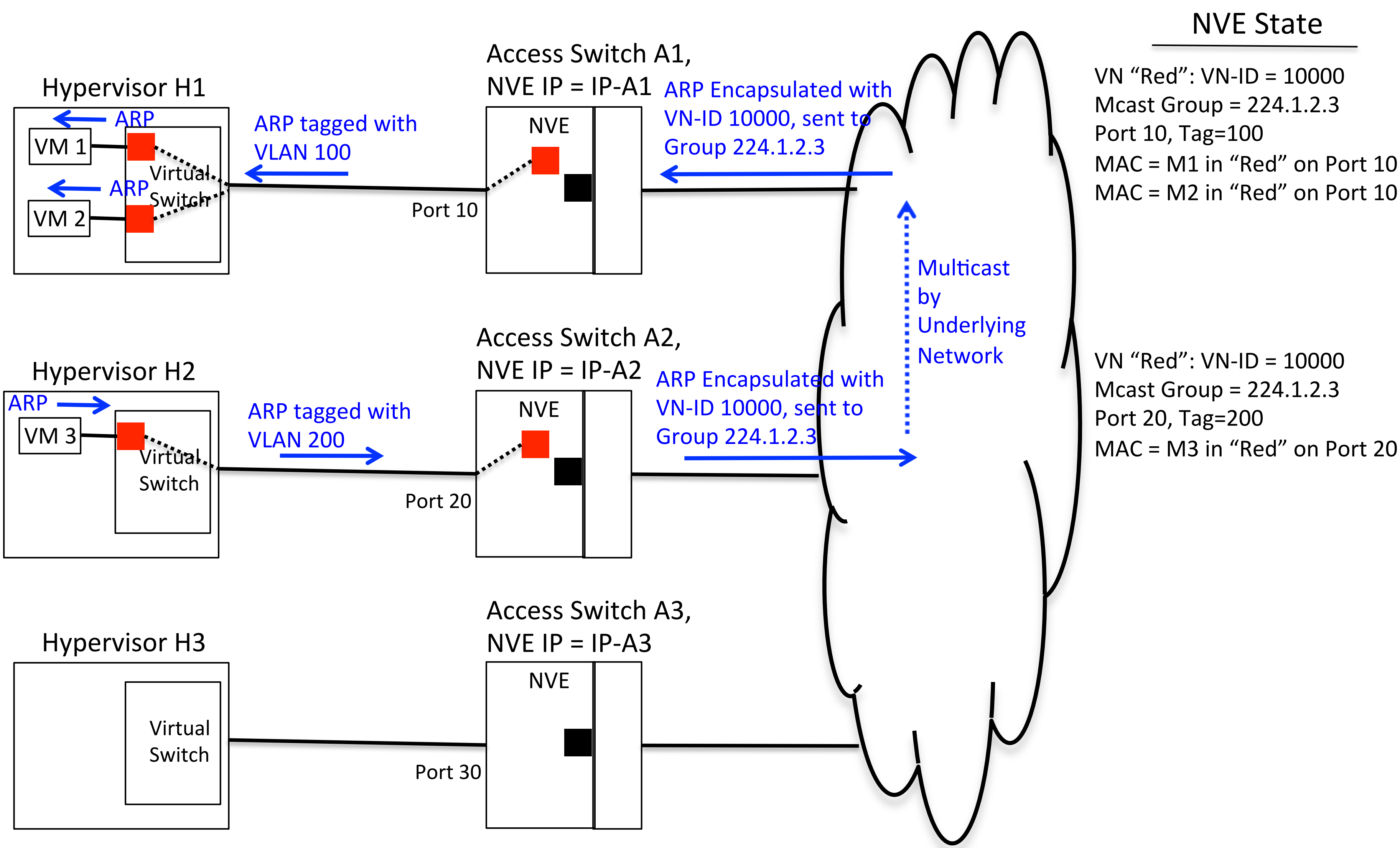
VM 3 comes up on Hypervisor H2, connected the VN “Red”

H2’s Virtual Switch signals to A2 that MAC M3 is connected to VN “Red”



VM 3 ARPs for VM1

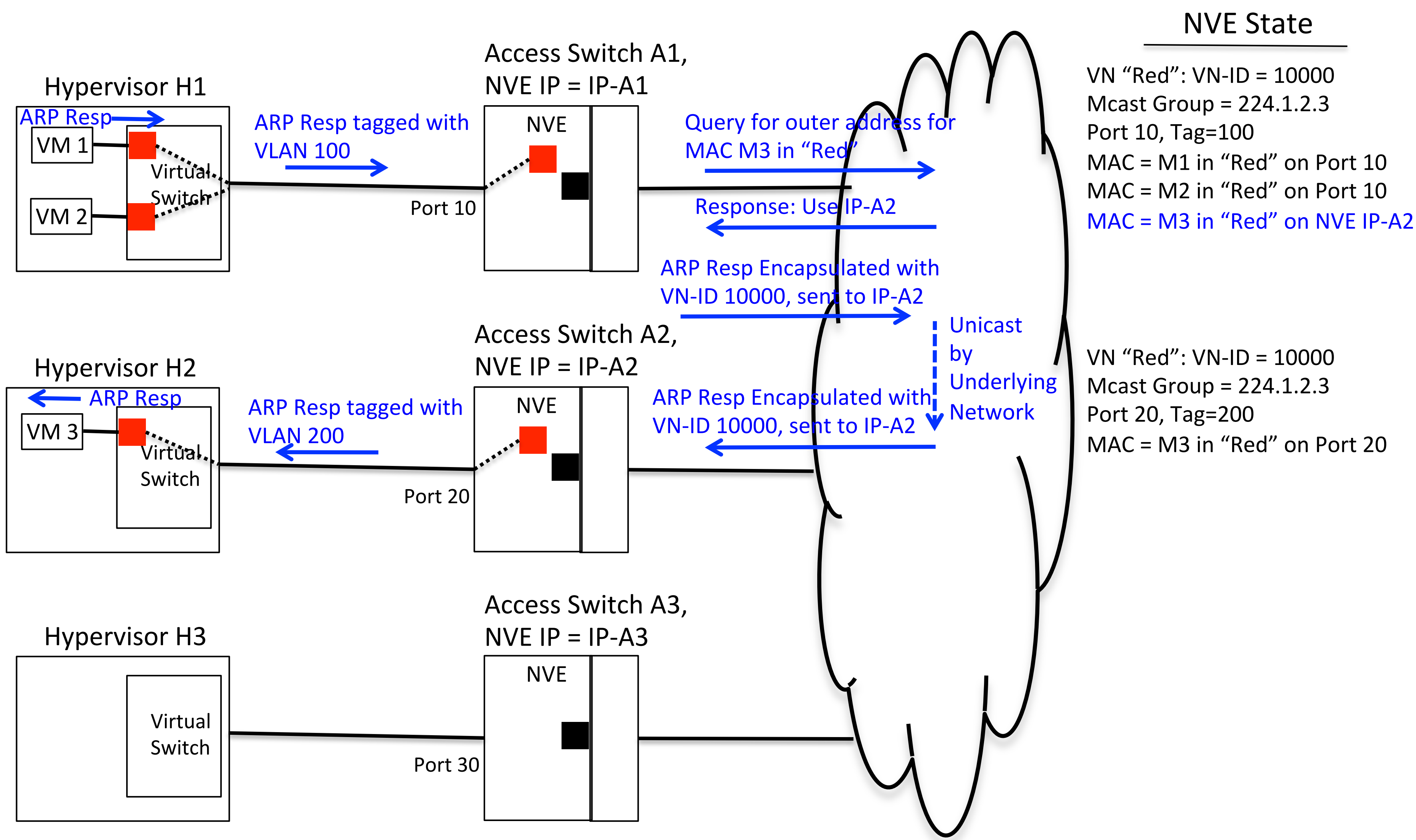
NVE A2 uses multicast to send the ARP Bcast to all NVEs interested in VN “Red”
NVE A1 Queries to find inner to outer mapping for MAC M3



VM 1 Sends ARP Response to VM3

NVE A1 Queries central entity to find inner to outer mapping for MAC M3

NVE A1 Unicasts ARP Response to A2



Summary of CP Characteristics

- Lightweight for NVE
  - This means:
    - Low amount of state (only what is needed at the time)
    - Low on complexity (keep it simply)
    - Low on overhead (don't drain resources from NVE)
    - Highly Scalable (don't collapse when scaled)
- Extensible
    - Support multiple address families (e.g. IPv4 and IPv6)
    - Allow addition of new address families
- Quickly reactive to change
    - Support Live Migration of VMs

Conclusion

- Two Categories of Control Plane protocols are needed to support a dynamic virtualized data center to dynamically build the state needed by an NVE to perform its map+encap and decap +deliver function.
- There are several models of operation possible which the WG will need to decide on.
- To help in deciding, the draft contains important evaluation criteria to use for comparing proposed solutions.