NVO3: Network Virtualization Problem Statement

Thomas Narten
narten@us.ibm.com

IETF 83 – Paris
March, 2012

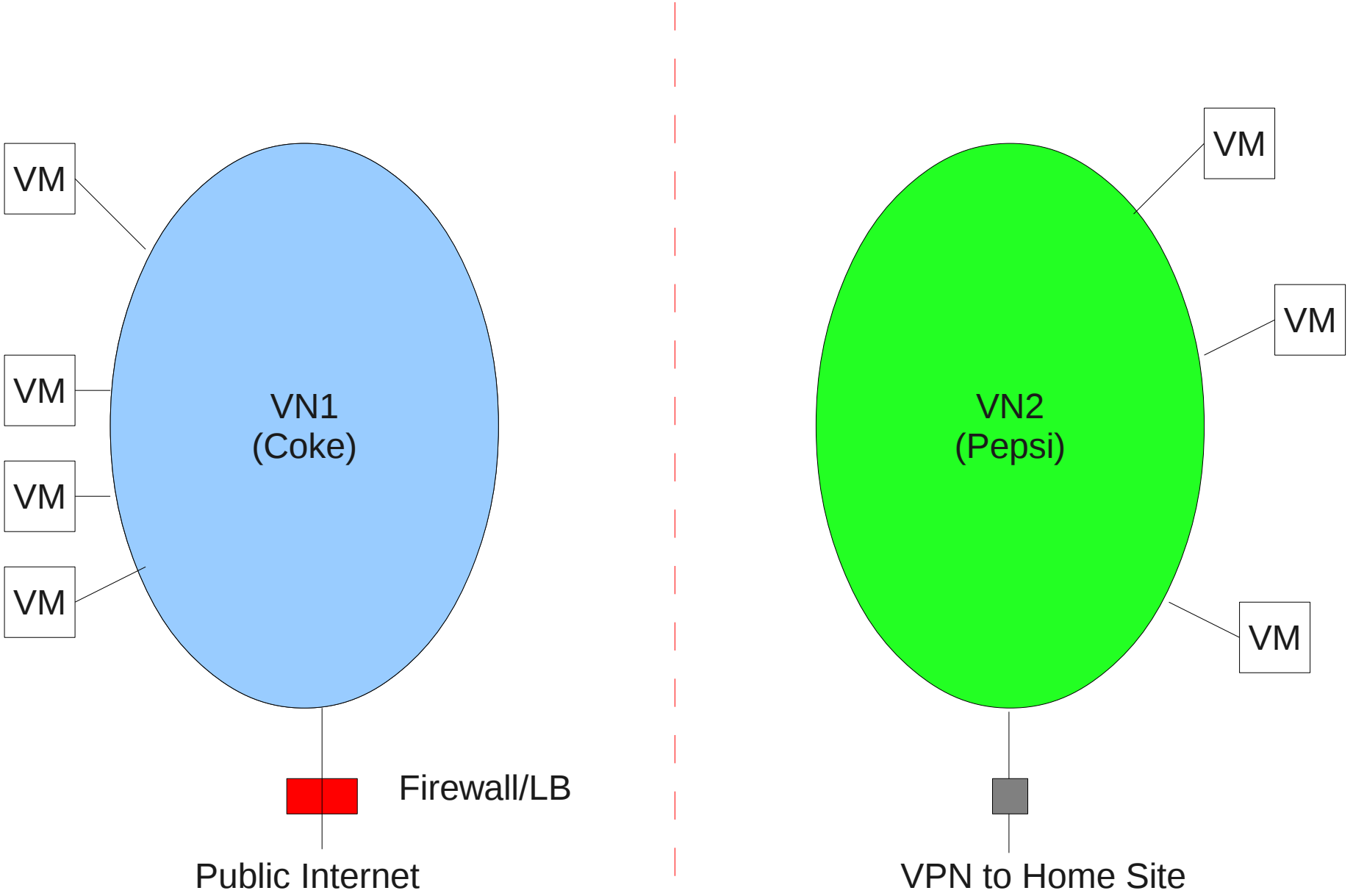
High-Level Motivation

- Imagine a data center
  - Could be cloud provider, hosting center, enterprise
  - Supports multiple tenants (e.g., Pepsi and Coke).
- Tenant wants (and operator wants to sell) ability to:
  - Create a Virtual Network instance
  - Logically attach a set of VMs (or machines) to the Virtual Network
  - Provide “Network as a Service”
- The Virtual Network (with associated VMs) provides a self-contained distributed service
  - E.g., web hosting, email service, etc.
  - May also use VPN to extend back into enterprise network

VN Requirements (Tenant Perspective)

- VMs think they are connected to a "real" network
  - Send/receive Ethernet frames
- Each VN instance uses its own IP address space
  - Tenant uses whatever IP addresses it wants (e.g., private addresses)
- VNs are isolated from each other (security)
  - One tenant's traffic not visible to other tenants
  - Ethernet frames stay local to a VN
  - Traffic forwarded to/from other networks only through controlled entry point(s)
    - E.g., connection to public Internet, VPN to tenant's home site, cross-VN IP router, etc.
    - Each entry point could include firewall, ACLs, etc.

Logical View (Tenant)



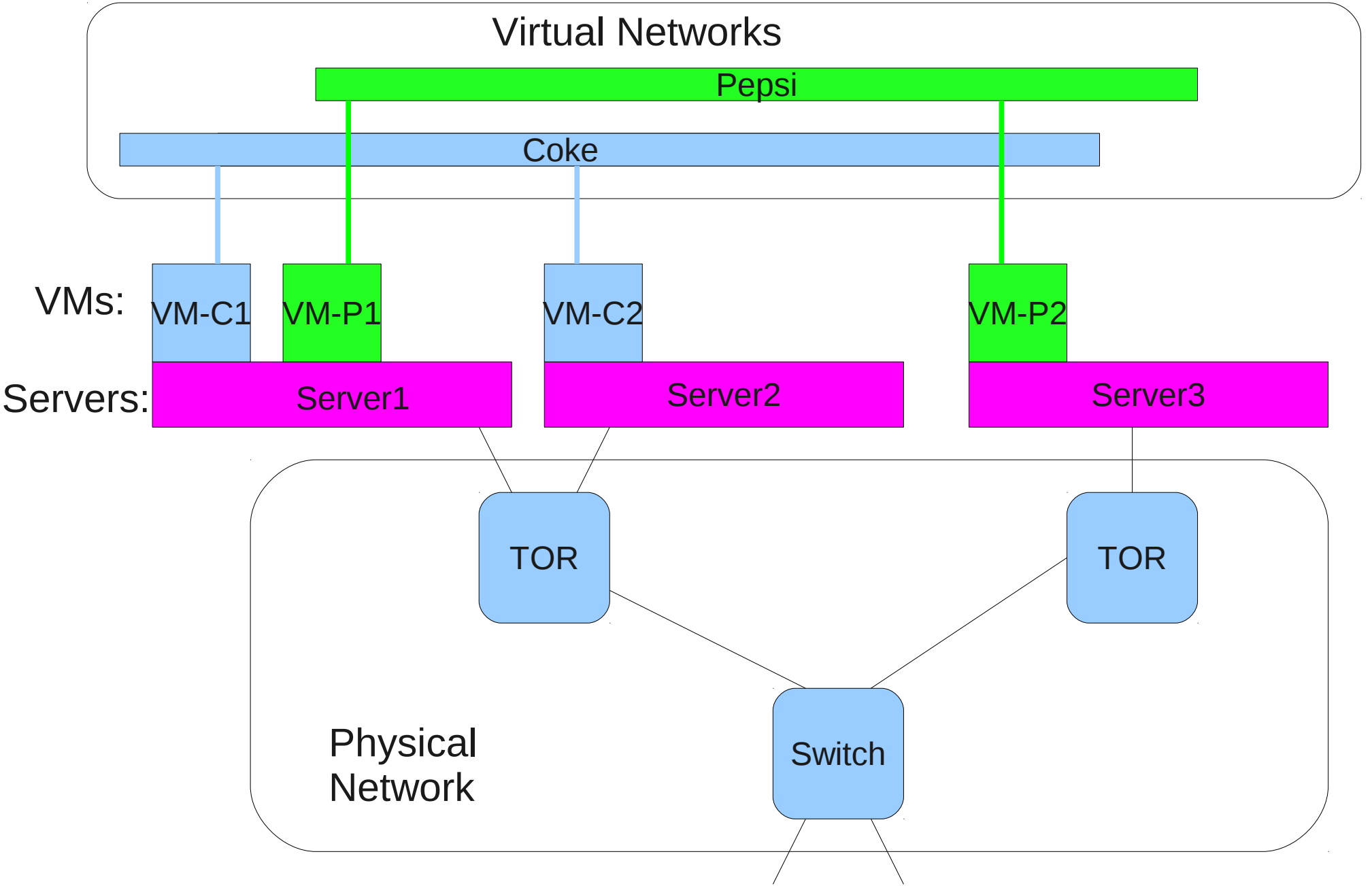
VN Requirements (DC Perspective)

- Want ability to place VMs anywhere within data center
  - Without being constrained by physical network attributes or concerns (e.g., IP subnet boundaries)
  - Both initial placement and for VM migration
- In practice, VMs can only be moved within an IP subnet
  - Big IP subnets imply large L2 Ethernet domains
- But, large L2 domains spanning entire DC increasingly painful
  - ARMD has been looking at address resolution issues
  - VLAN space limitations are a well-recognized problem
  - Large L2 domains increase fate sharing concerns
  - Issues will only get worse in future as size of DCs grow
  - TRILL, SPB, etc. working on this at L2, but no magic bullet
- Note: Above two are in conflict with each other

Requirements (DC Perspective) – Cont

- Want to separate the logical network attributes associated with VM from the physical instantiation
  - e.g, VLAN info, QoS, L2 protocols, IP Subnets, etc.
  - Observation: reconfiguring the network elements when placing VMs is complex, error prone
- Want to abstract away the key network properties
  - Server virtualization allows VMs to abstract away physical properties for memory, processor, I/O, etc.
  - Network properties include VLANs, IP Subnetting, etc.
- Solution needs to scale to cover entire data center (and beyond)
  - Millions of VMs (and beyond)

Physical & Logical View



DC Multi-Tenancy Today

- Implemented at L2 today (Ethernet VLANs)
- But, increasingly painful to scale as size of DC increases
  - VLAN limitation
  - Size of forwarding tables in core switches is an issue
  - Increased need for multipathing
  - Large IP Subnets/L2 domains better support VM migration, but smaller L2 domains better for fate-sharing containment
- TRILL, SPB, etc. will help, but are L2 solutions
  - Network service provided remains Ethernet
  - No magic bullet

L3 Overlay Multi-Tenancy Approach

- Increase use of IP within DC core, reuse proven IP technology:
  - ECMP load balancing, etc.
  - Fast routing convergence (OSPF, ISIS, etc.)
  - Rich ecosystem of existing technology/products
- Incremental deployability with existing DC network
  - Deploy at hypervisor or edge switch, few or no changes to core network
- L3 overlays offer compelling value proposition
  - In fully virtualized systems, can be implemented entirely in hypervisor software, without network changes
  - In traditional DCNs, enable edge switches, no change to rest of network
  - TRILL, SPB, etc. have significantly different deployment paths

Summary of Requirements

- Multi-tenant support, scaling to the millions of VMs
- Support VM placement anywhere in data center
  - Both initial placement & migration
- DC-driven requirements
  - More reliance on IP, less on large L2 domains
  - Incremental deployability in existing DC network
  - Strong hesitancy to deploy BGP/MPLS VPN technologies into DC
- On-demand elastic provisioning of resources
  - Grow/shrink dynamically as workload changes
- Decouple logical network configuration from physical instantiation

Acknowledgments

This effort stems from the work of many others...

On problem statement document: Murari Sridharan, Dinesh Dutt, David Black, Lawrence Kreeger.

See list of authors in the NVGRE and VXLAN documents....