Network Fundamentals for Cloud

BITS Pilani
Pilani Campus

Nishit Narang
WILPD-CSIS

innovate achieve lead

BITS Pilani
Pilani Campus

CC ZG503: Network Fundamentals for Cloud
Lecture No. 8

innovate achieve lead

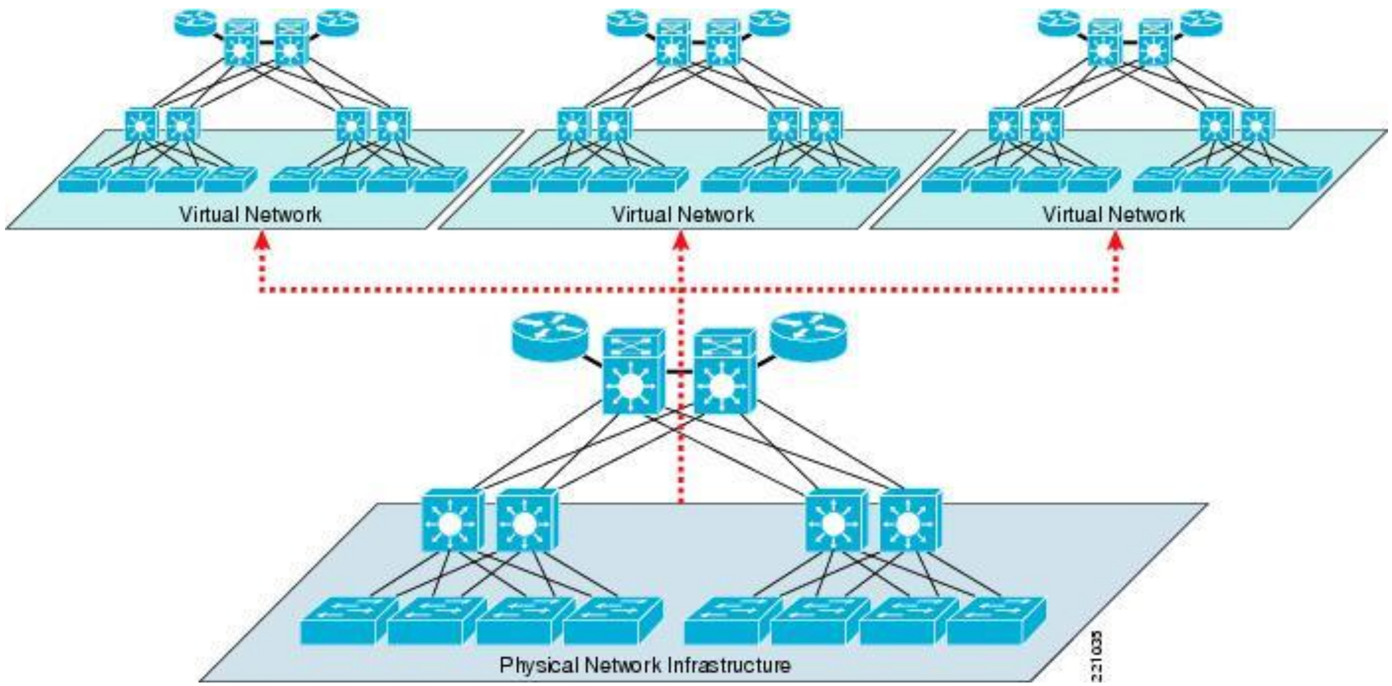
BITS Pilani
Pilani Campus

Enabling Technologies: NV/NFV (Contd.)

Network Virtualization and Network Function Virtualization

RECAP: Network Virtualization

- What is network virtualization ?

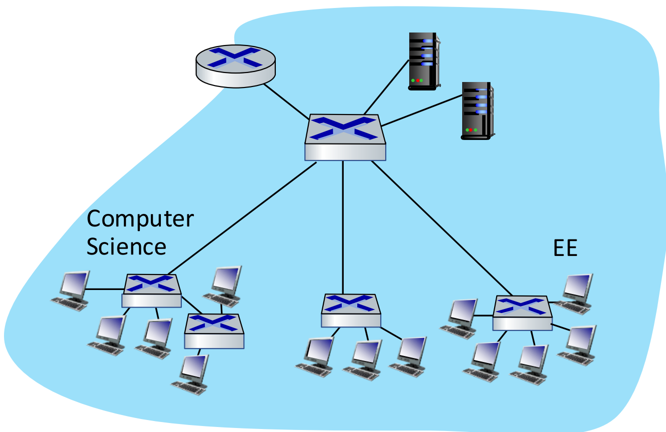


RECAP: External Network Virtualization

- External network virtualization in different layers :
  - Layer 1
    - Seldomly implemented in physical data transmission layer.
  - Layer 2
    - Use some tags in MAC address packet to provide virtualization.
    - Example, VLAN.
  - Layer 3
    - Use some tunnel techniques to form a virtual network.
    - Example, VPN.
  - Layer 4 or higher
    - Build up some overlay network for some application.
    - Example, P2P.

Virtual LANs (VLANs): motivation

Q: what happens as LAN sizes scale, users change point of attachment?

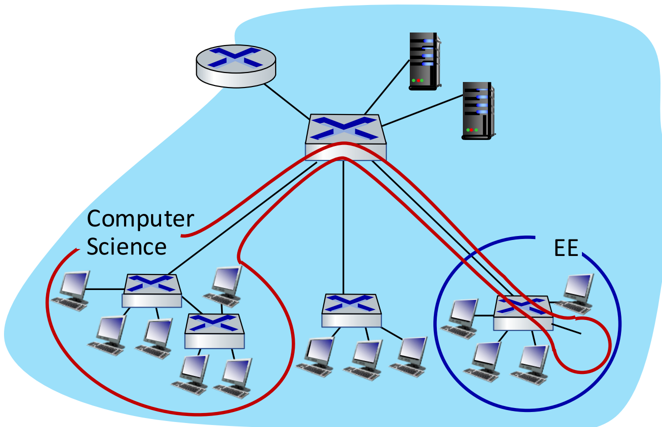


single broadcast domain:

- *scaling:* all layer-2 broadcast traffic (ARP, DHCP, unknown MAC) must cross entire LAN
- efficiency, security, privacy issues

Virtual LANs (VLANs): motivation

Q: what happens as LAN sizes scale, users change point of attachment?



single broadcast domain:

- *scaling:* all layer-2 broadcast traffic (ARP, DHCP, unknown MAC) must cross entire LAN
- efficiency, security, privacy, efficiency issues

administrative issues:

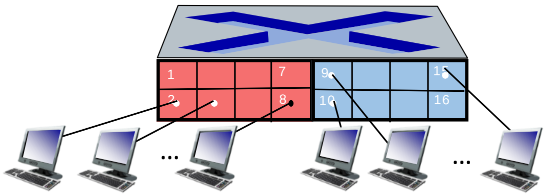
- CS user moves office to EE - *physically* attached to EE switch, but wants to remain *logically* attached to CS switch

Port-based VLANs

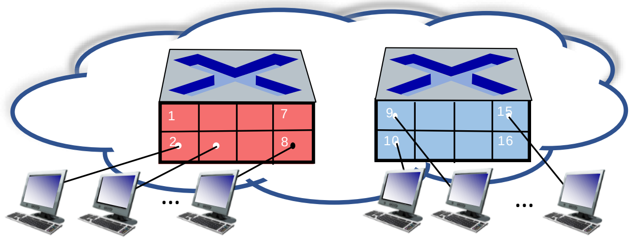
Virtual Local Area Network (VLAN)

switch(es) supporting VLAN capabilities can be configured to define multiple *virtual* LANS over single physical LAN infrastructure.

port-based VLAN: switch ports grouped (by switch management software) so that *single* physical switch ……

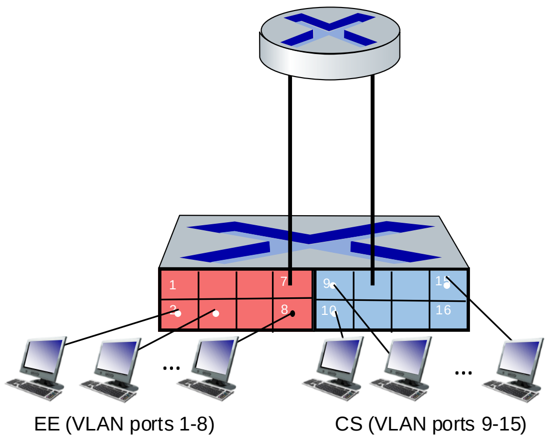


… operates as multiple virtual switches

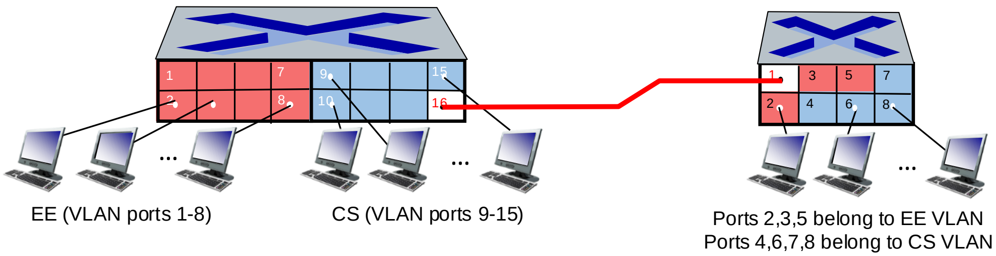


Port-based VLANs

- traffic isolation: frames to/from ports 1-8 can *only* reach ports 1-8
  - can also define VLAN based on MAC addresses of endpoints, rather than switch port
- dynamic membership: ports can be dynamically assigned among VLANs
- forwarding between VLANS: done via routing (just as with separate switches)
  - in practice vendors sell combined switches plus routers



VLANS spanning multiple switches



trunk port: carries frames between VLANS defined over multiple physical switches

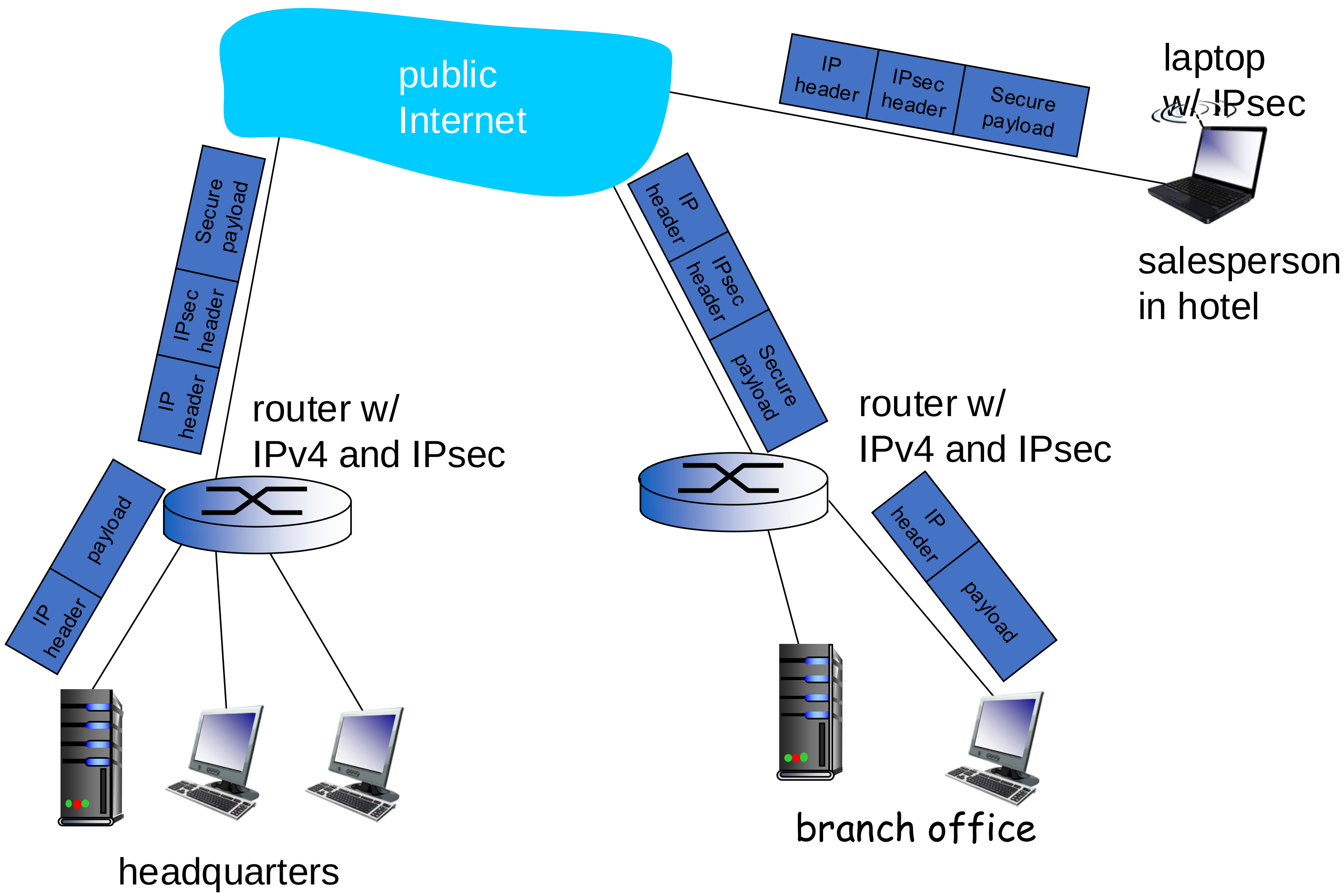
- frames forwarded within VLAN between switches can't be vanilla 802.1 frames (must carry VLAN ID info)
- 802.1q protocol adds/removed additional header fields for frames forwarded between trunk ports

Virtual Private Networks (VPNs)

motivation:

- institutions often want private networks for security.
  - costly: separate routers, links, DNS infrastructure.
- VPN: institution's inter-office traffic is sent over public Internet instead
  - encrypted before entering public Internet
  - logically separate from other traffic

Virtual Private Networks (VPNs)



IPsec services

- data integrity
- origin authentication
- replay attack prevention
- confidentiality

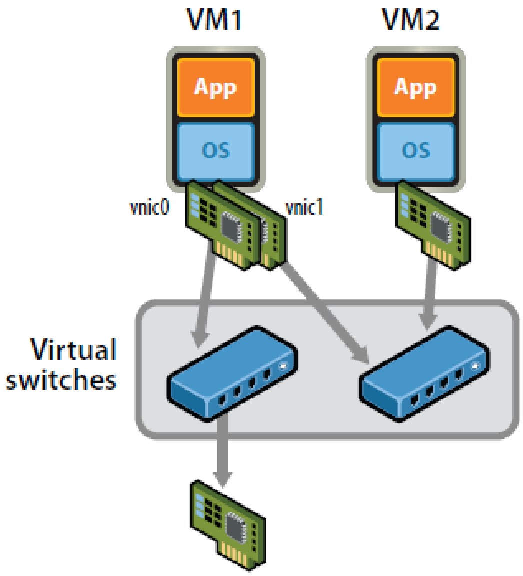
- two protocols providing different service models:
  - AH
  - ESP

Internal Network Virtualization

- Internal network virtualization
  - A single system is configured with containers, such as the Xen domain, combined with hypervisor control programs or pseudo-interfaces such as the VNIC, to create a “network in a box”.
  - This solution improves overall efficiency of a single system by isolating applications to separate containers and/or pseudo interfaces.
  - Virtual machine and virtual switch :
    - The VMs are connected logically to each other so that they can send data to and receive data from each other.
    - Each virtual network is serviced by a single virtual switch.
    - A virtual network can be connected to a physical network by associating one or more network adapters (uplink adapters) with the virtual switch.

Internal Network Virtualization

- Properties of virtual switch
  - A virtual switch works much like a physical Ethernet switch.
  - It detects which VMs are logically connected to each of its virtual ports and uses that information to forward traffic to the correct virtual machines.
- Typical virtual network configuration
  - Communication network
    - Connect VMs on different hosts
  - Storage network
    - Connect VMs to remote storage system
  - Management network
    - Individual links for system administration



Network Functions Virtualization

- Network Functions Virtualization is about implementing network functions in software - that today run on proprietary hardware - leveraging (high volume) standard servers and IT virtualization
- Supports multi-versioning and multi-tenancy of network functions, which allows use of a single physical platform for different applications, users and tenants
- Enables new ways to implement resilience, service assurance, test and diagnostics and security surveillance

Network Functions Virtualization

- Provides opportunities for pure software players
- Facilitates innovation towards new network functions and services that are only practical in a pure software network environment
- Applicable to any data plane packet processing and control plane functions, in fixed or mobile networks
- NFV will only scale if management and configuration of functions can be automated
- NFV aims to ultimately transform the way network operators architect and operate their networks, but change can be incremental

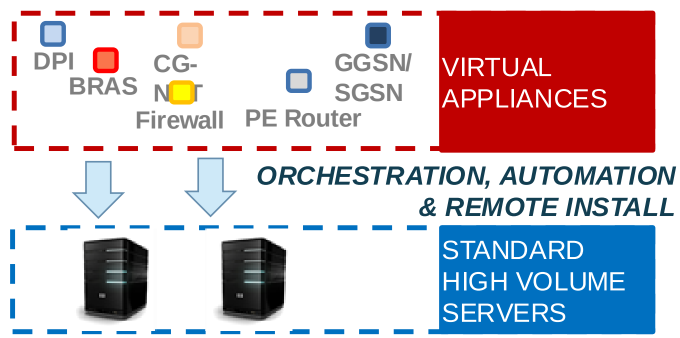
The NFV Concept

A means to make the network more flexible and simple by minimising dependence on HW constraints

Traditional Network Model: APPLIANCE APPROACH

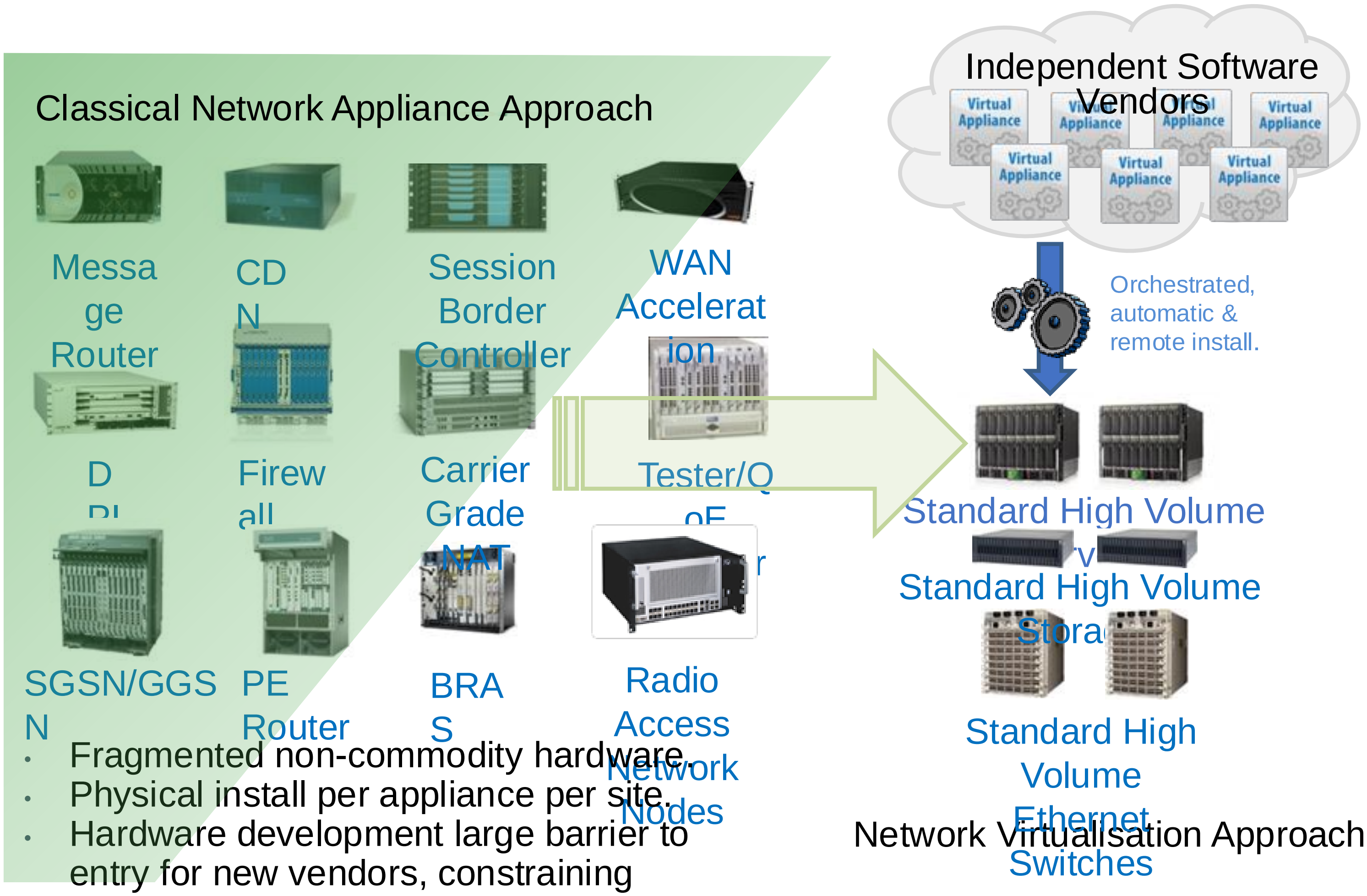
- Network Functions are **based on specific HW&SW**
- **One physical node per role**

Virtualised Network Model: VIRTUAL APPLIANCE APPROACH



- Network Functions are **SW-based over well-known HW**
- **Multiple roles** over **same HW**

Target



- Fragmented non-commodity hardware.
- Physical install per appliance per site.
- Hardware development large barrier to entry for new vendors, constraining innovation & competition.

Benefits & Promises of NFV

- Flexibility to easily, rapidly, dynamically provision and instantiate new services in various locations
- Improved operational efficiency
  - by taking advantage of the higher uniformity of the physical network platform and its homogeneity to other support platforms.
- Software-oriented innovation to rapidly prototype and test new services and generate new revenue streams
- More service differentiation & customization
- Reduced (OPEX) operational costs: reduced power, reduced space, improved network monitoring
- IT-oriented skillset and talent

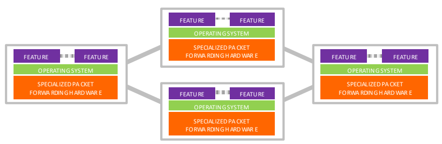
NFV vs SDN

- NFV: re-definition of **network equipment architecture**
- NFV was born to meet Service Provider (SP) needs:
  - Lower CAPEX by reducing/eliminating proprietary hardware
  - Consolidate multiple network functions onto industry standard platforms
- SDN: re-definition of **network architecture**
- SDN comes from the IT world:
  - Separate the data and control layers, while centralizing the control
  - Deliver the ability to program network behavior using well-defined interfaces

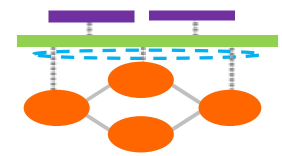
Software Defined Networking

Network equipment as
Black boxes

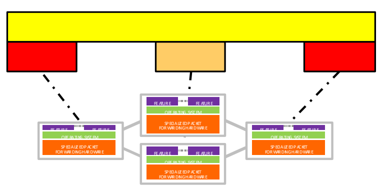
Open interfaces (OpenFlow)
for instructing the boxes
what to do



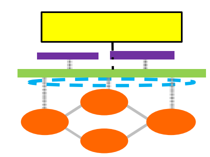
Boxes with autonomous
behaviour



Decisions are taken out of the box

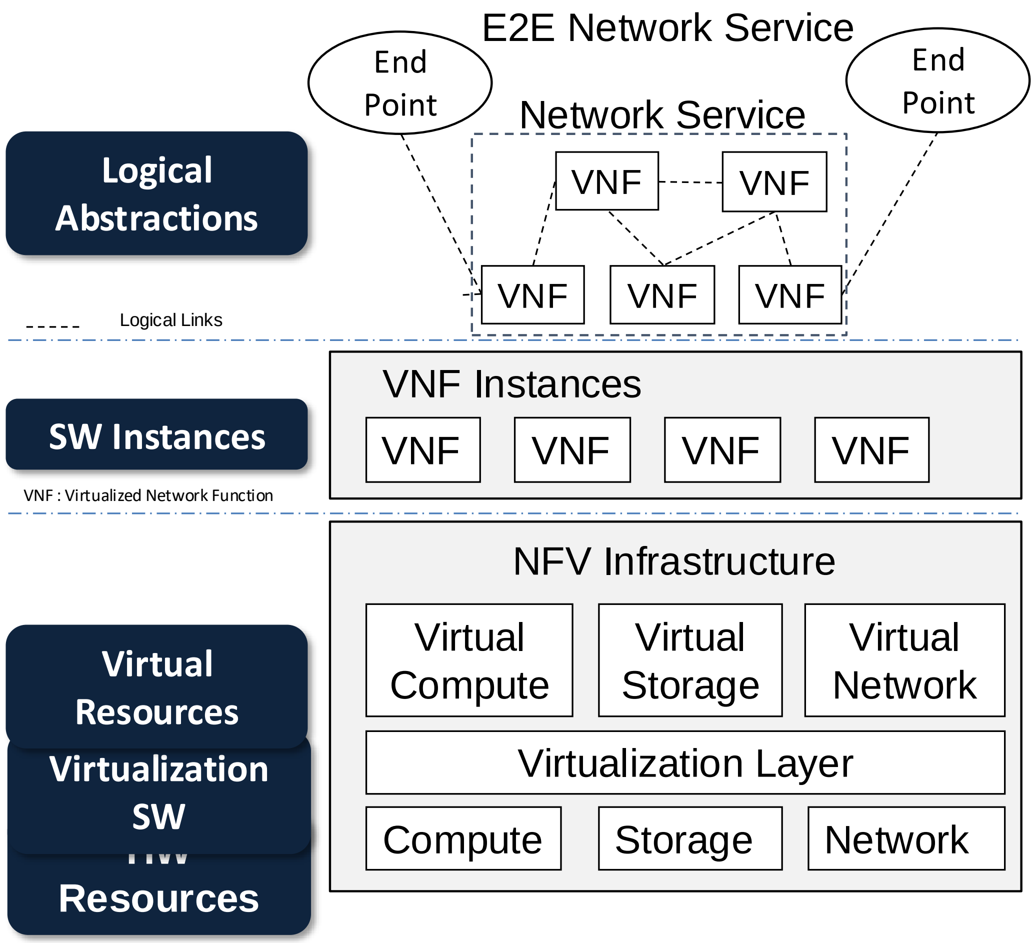


Adapting OSS to manage black boxes



Simpler OSS to manage the SDN
controller

NFV Layers



E2E Network Service
End Point
Network Service
End Point
Logical Abstractions
VNF
VNF
VNF
VNF
VNF
Logical Links
VNF Instances
SW Instances
VNF
VNF
VNF
VNF
VNF : Virtualized Network Function
NFV Infrastructure
Virtual Resources
Virtual Compute
Virtual Storage
Virtual Network
Virtualization SW
Virtualization Layer
HW Resources
Compute
Storage
Network

SDN/NFV in the Data Center

- NFV Data Center
  - Used by service providers to host communications and networking services
  - Services can be loaded as cloud-based software on commercial off-the-shelf (COTS) server hardware
  - Applications are hosted in data center so they could be accessed via cloud

- SDN can work in tandem with NFV
  - Traffic Steering in an NFV Data Center

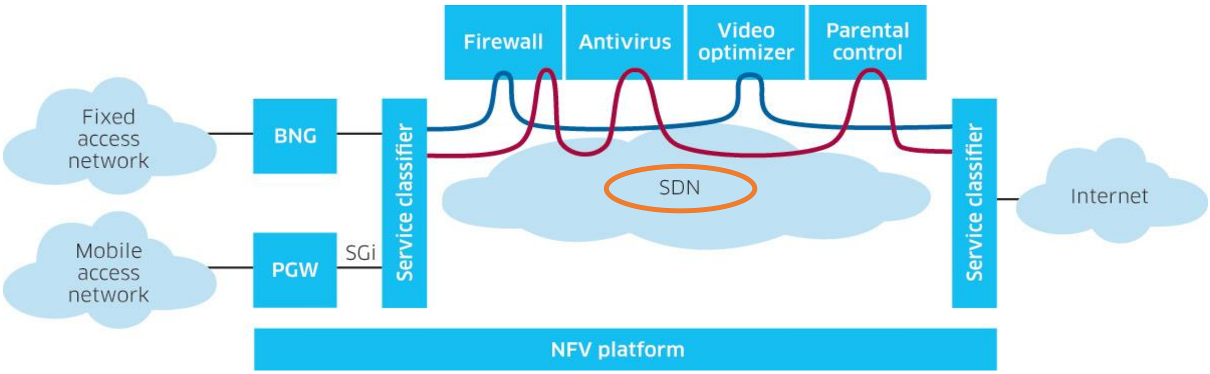


Image Source: SDxCentral

innovate achieve lead

BITS Pilani
Pilani Campus

CC ZG503: Network Fundamentals for Cloud

Data Center Network (DCN)

Networks in the Cloud

- Cloud Networking
  - Physical ***vs*** Virtual
  - Underlay ***vs*** Overlay
  - Intra- ***vs*** Inter-

- Data Center Networks studies a specific aspect of Cloud Networking

Course Structure

Network Performance and Security

Virtual Cloud Networks

Data Center Networks
(Physical Network)

Enabling Technology Overview

Network Fundamentals

Role of DCN

- Role of DC in realizing the “Cloud”
  - DC provides the ingredients for the Cloud
    - Compute (of different capacities & types)
    - Storage (of different capacities & types)
- Networking “Connects” the ingredients!!

- DCN → Data Center Network
  - The network that connects the assets within a Data Center

- Video
  - https://www.youtube.com/watch?v=avP5d16wEp0

DCN vs Cloud DCN??

- Cloud DCN = a new type of DC based on cloud computing architecture
- Within a cloud DC, various IT devices are fully virtualized
- A cloud DC features virtualized servers, storage devices, and applications, enabling users to leverage various resources on demand

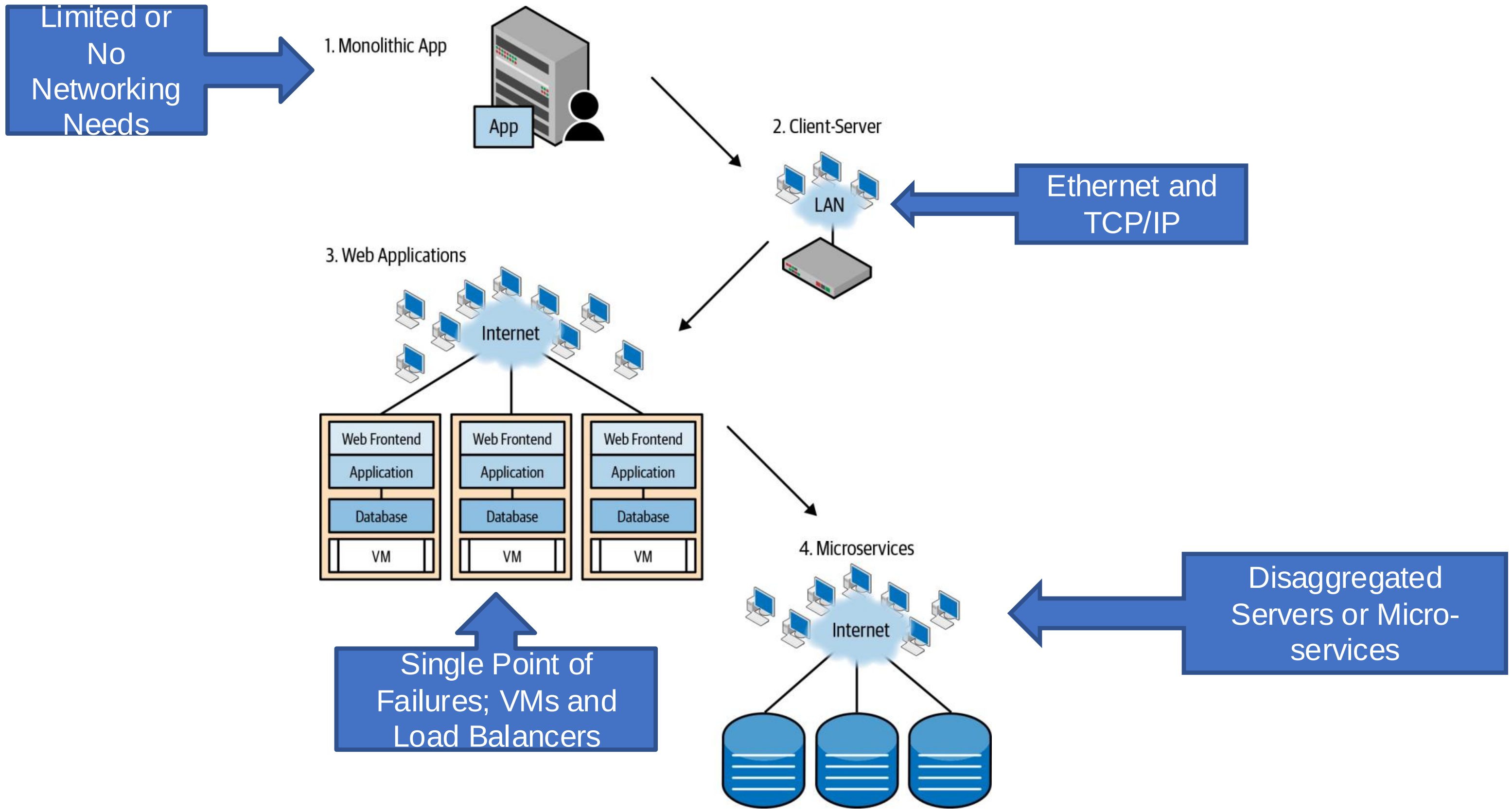
Application - Network Dance!

"Once upon a time, there was what there was, and if nothing had happened there would be nothing to tell" — Charles de Lint (**Canadian writer of Dutch,** Spanish, and Japanese ancestry)

- A distributed application is in a dance with the network, with the application leading

- The story of the modern data center network begins when the network was caught flat-footed when the application began the dance to a different tune

Evolution of the Application



Cluster-based application architectures such as MapReduce have become prominent

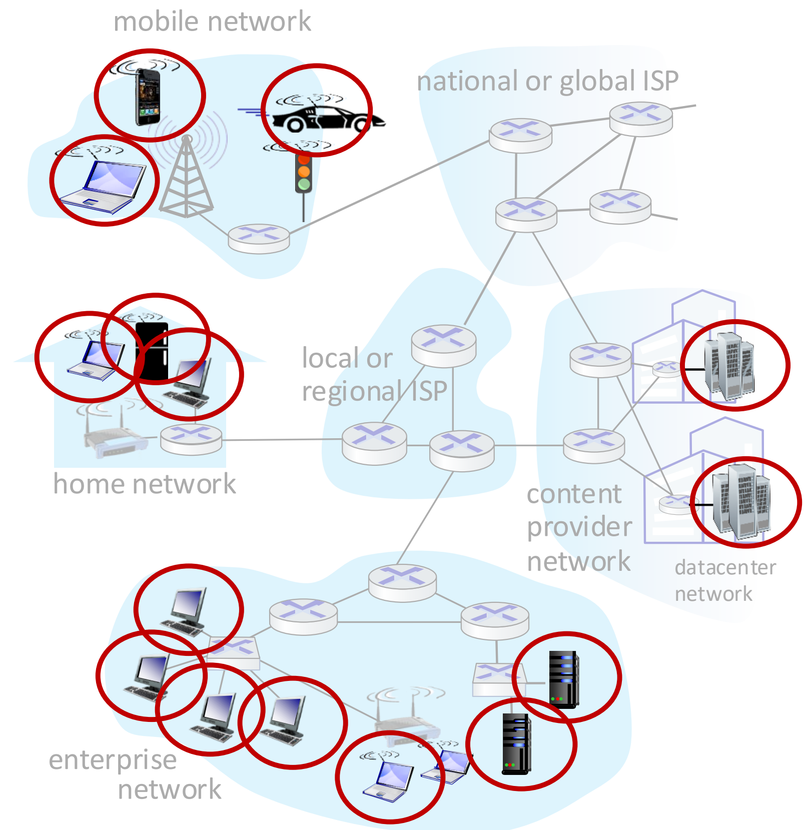
Historic Shift from Client-Server Traffic to Server-Server Traffic

Source: [Cloud Native Data Center Networking by Dinesh G. Dutt]

A closer look at Internet structure

Network edge:

- hosts: clients and servers
- servers often in data centers



Source: Computer Networking: A Top-Down Approach, 8th edition, Jim Kurose, Keith Ross, Pearson, 2020

Facebook Traffic

- Case Study
  - [Introducing data center fabric, the next-generation Facebook data center network - Engineering at Meta]

"Facebook's network infrastructure needs to constantly scale and evolve, rapidly adapting to our application needs. The amount of traffic from Facebook to Internet – we call it "machine to user" traffic – is large and ever increasing, as more people connect and as we create new products and services. However, this type of traffic is only the tip of the iceberg. What happens inside the Facebook data centers – "machine to machine" traffic – is several orders of magnitude larger than what goes out to the Internet."

DCN Applications and Traffic Patterns

- Cluster-based Applications
  - Important application classes include scientific computing, financial analysis, data analysis and warehousing
- Principle bottleneck in large-scale clusters is often inter-node communication bandwidth
- Examples:
  - MapReduce must perform significant data shuffling to transport the output of its map phase before proceeding with its reduce phase
  - Applications running on cluster-based file systems often require remote-node access before proceeding with their I/O operations
  - A query to a web search engine often requires parallel communication with every node in the cluster hosting the inverted index to return the most relevant results
  - Retrieval of a single web page can require coordination and communication with literally hundreds of individual sub-services running on remote nodes

Reference Source: A Scalable, Commodity Data Center Network Architecture by Al-Fares et al, Department of Computer Science and Engineering University of California, San Diego La Jolla, CA

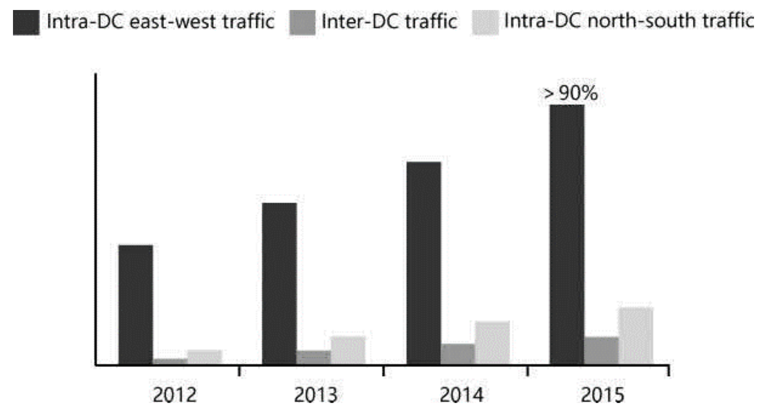
DCN Challenges

- Internal Traffic is BIG
- Deadlines for Network I/O are TIGHT
- Congestion and TCP Incast

These aspects are discussed in the following slides.

DCN Challenges - Traffic

- Internal Traffic is BIG!!
  - In 2020, every minute there were
    - more than 1.6 million Google searches,
    - 260 million emails sent,
    - 47,000 apps downloaded,
    - 220,000 photos uploaded to Facebook, and
    - 660 million data packets transmitted…..
  - It is estimated that global DC IP traffic increases five-fold every year



- Big data requires wide pipes…..
  - East-West traffic accounts for more than 90% of total DC traffic

Source: Lei Zhang, Le Chen. Cloud Data Center Network Architectures and Technologies, CRC Press 2021

DCN Challenges – Network Faults and Capacity

- Need for Intelligent O&M and Network Fault Recovery
  - Driven by cloud-based DCs and Network Function Virtualization (NFV), the number of managed objects (MOs) in a cloud DCN is ten times greater than that of a legacy DCN
  - Network needs to detect dynamic VM migration and elastic scaling of applications, which results in frequent configuration changes and traffic surges
    - Example: LinkedIn data shows that the number of network faults saw an 18-fold increase from 2010 to 2015.
  - As network, computing, and storage boundaries are blurred, network faults become more difficult to locate and isolate
- Types of Network Faults:
  - Connection faults, such as a VM going offline unexpectedly or communication becoming intermittently interrupted.
  - Performance faults, such as network congestion during heavy loads.
  - Policy faults, such as unauthorized access and port scanning.
- Newer applications are demanding tight Network-I/O timings!
  - Also see next slide….

DCN Challenges – Network Faults and Capacity

- How to address Network Faults and Capacity Challenges?
  - Use of Self-healing networks
  - Use of Intelligent analysis engine (big data algorithms) to predict / detect / isolate network faults
  - Use of SDN and SDN Controller for simplified cloud network operations
  - Bandwidth Oversubscription (especially in core layer)

innovate achieve lead

DCN Challenges – TCP Incast

- TCP incast is a recently identified network transport pathology that affects many-to-one communication patterns in datacenters
- Caused by a complex interplay between datacenter applications, the underlying switches, network topology, and TCP, which was originally designed for wide area networks

- The problem especially affects computing paradigms in which distributed processing cannot progress until all parallel threads in a stage complete
- Examples of such paradigms include distributed file systems, web search, advertisement selection, and other applications with partition or aggregation semantics

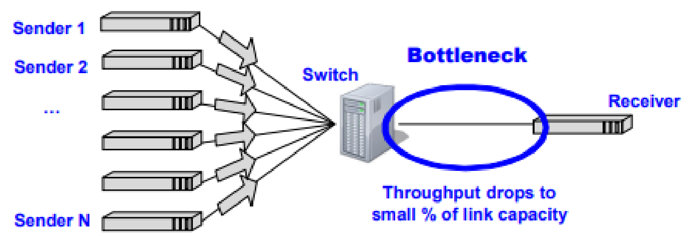


Figure 1. Simple setup to observe incast. The receiver requests k blocks of data from a set of N storage servers. Each block is striped across N storage servers. For each block request received, a server responds with a fixed amount of data. Clients do not request block $k+1$ until all the fragments of block k have been received.

Source: Understanding TCP Incast and Its Implications for Big Data Workloads by Yanpei Chen, Rean Griffith*, David Zats, Anthony D. Joseph, Randy Katz (University of California, Berkeley, *Vmware)

DCN Challenges – TCP Incast

- There have been many proposed solutions for TCP incast. Approaches include:
  - modifying TCP parameters or its congestion control algorithm,
  - optimizing application level data transfer patterns,
  - switch level modifications such as larger buffers or explicit congestion notification (ECN) capabilities, and
  - link layer mechanisms such as Ethernet congestion control.
- Application level solutions are the least intrusive to deploy, but require modifying each and every datacenter application
- Switch and link level solutions require modifying the underlying datacenter infrastructure, and are likely to be logistically feasible only during hardware upgrades
- TCP incast is fundamentally a transport layer problem, thus a solution at this level may be best
  - e.g. An existing solution is reducing the minimum length of TCP retransmission time out (RTO) from 200ms to 1ms

Source: Understanding TCP Incast and Its Implications for Big Data Workloads by Yanpei Chen, Rean Griffith*, David Zats, Anthony D. Joseph, Randy Katz (University of California, Berkeley, *Vmware)

DCN Challenges - Infrastructure

- DC Network Cost
- DC Cooling
- DC Cabling