# Cloud Networking (VITMMA02)
# Networks in the cloud

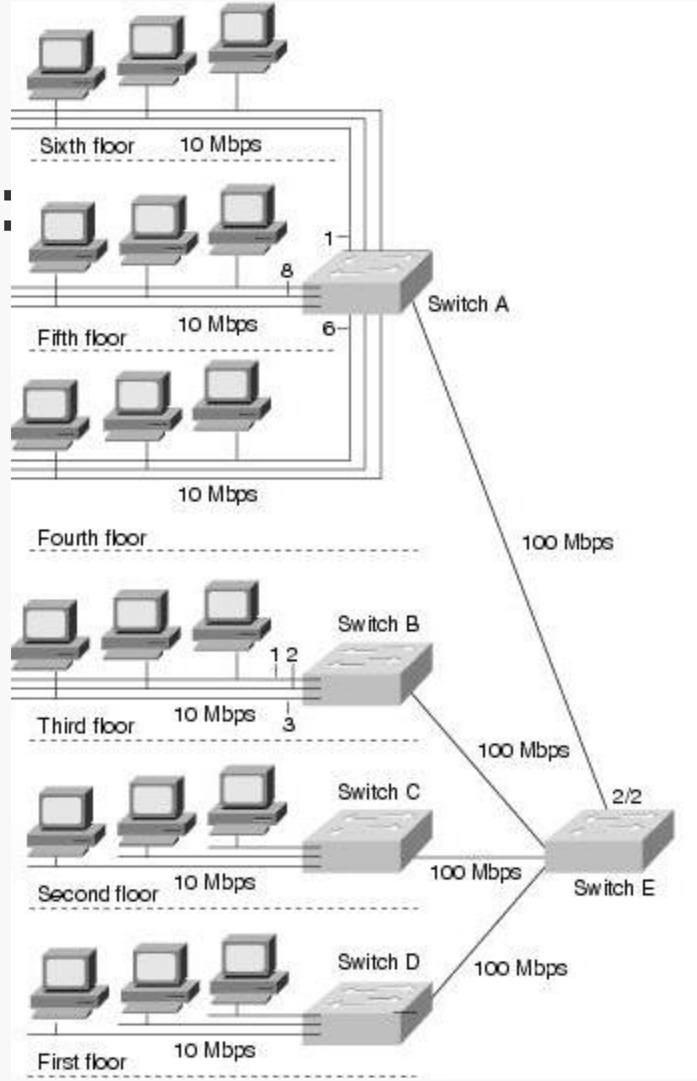
Markosz Maliosz PhD

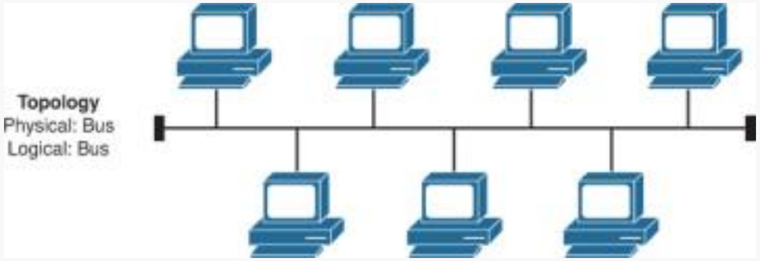
Department of Telecommunications and Media Informatics
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics

Spring 2019

# Ethernet

- Layer2 network
- Ethernet bridging or switching: bridged/switched Ethernet
  - emulating shared media

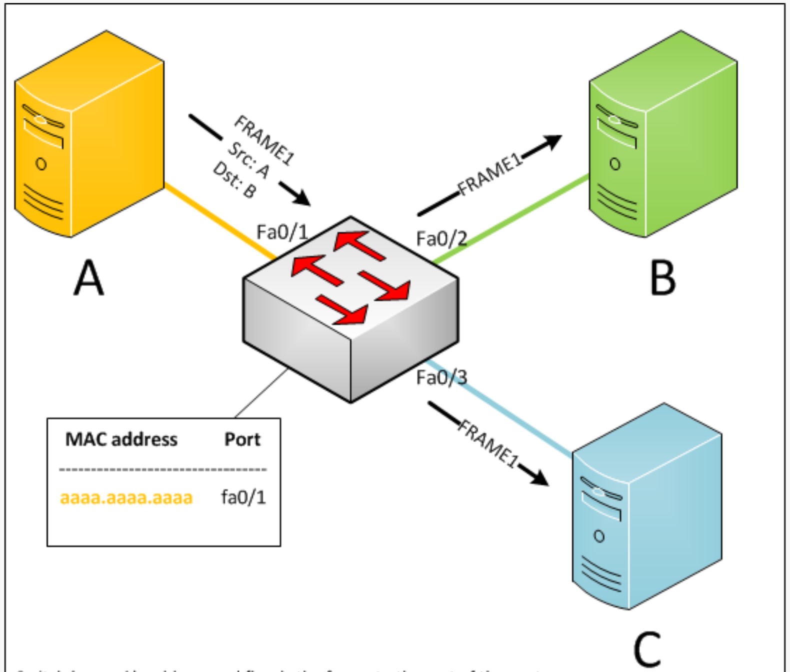




Backgrounder: http://www6.ietf.org/edu/documents/82-RoutingBridgingSwitching-Perlman.pdf pp.18-44

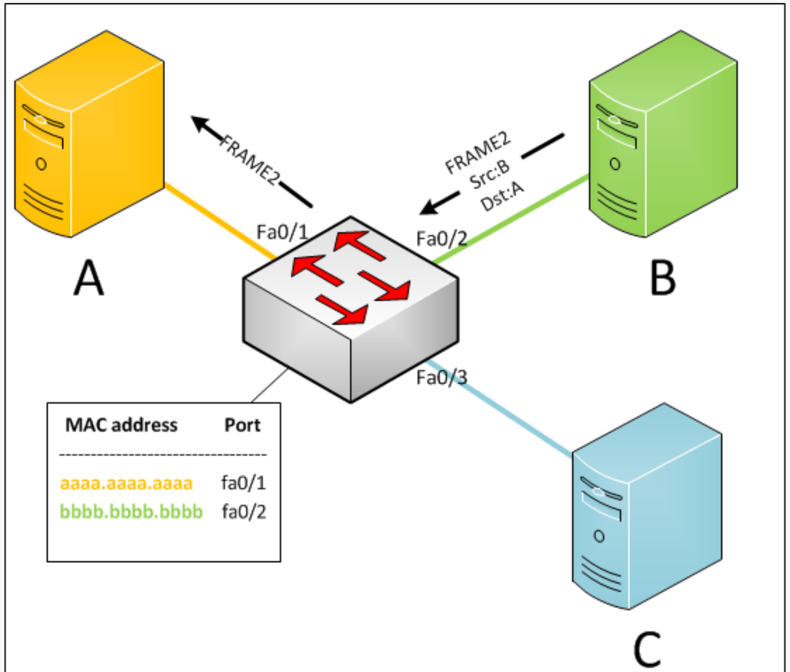
# Ethernet

- Spanning Tree Protocol (STP)
- MAC address learning
- transparent bridging
- flooding: broadcast, unknown unicast and multicast packets
- possible errors: implementation error, misconfigration
- forwarding loop is causing 100% CPU load for the affected switches



| MAC address | Port |
|---|---|
| aaaa.aaaa.aaaa | fa0/1 |

Switch learns A's address and floods the frame to the rest of the ports

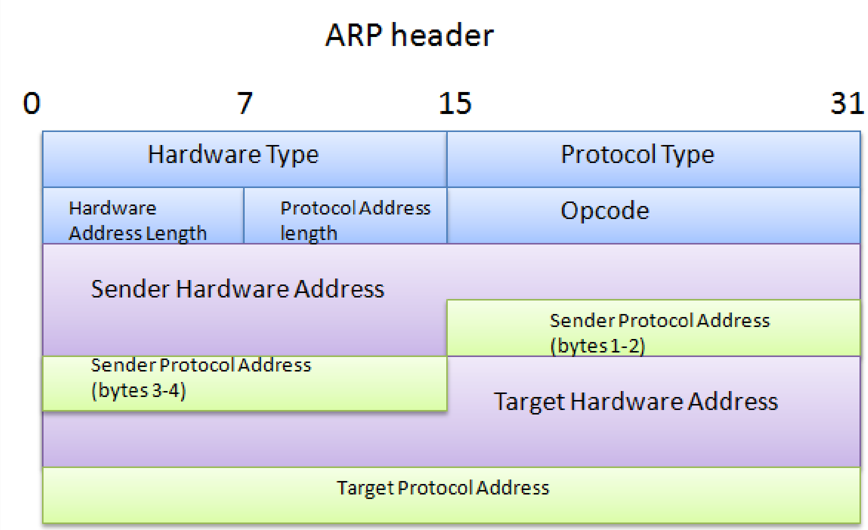


| MAC address | Port |
|---|---|
| aaaa.aaaa.aaaa | fa0/1 |
| bbbb.bbbb.bbbb | fa0/2 |

Switch learns B's address and forwards the frame to A

# IP address and MAC address mapping

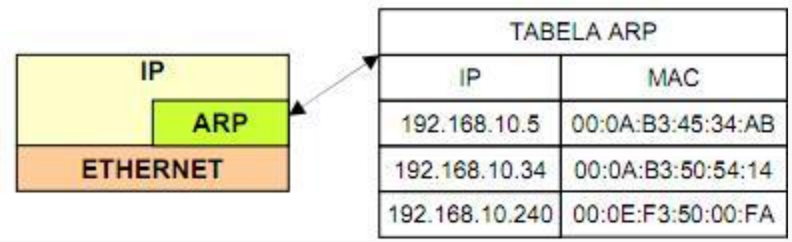
» Address Resolution Protocol (ARP)



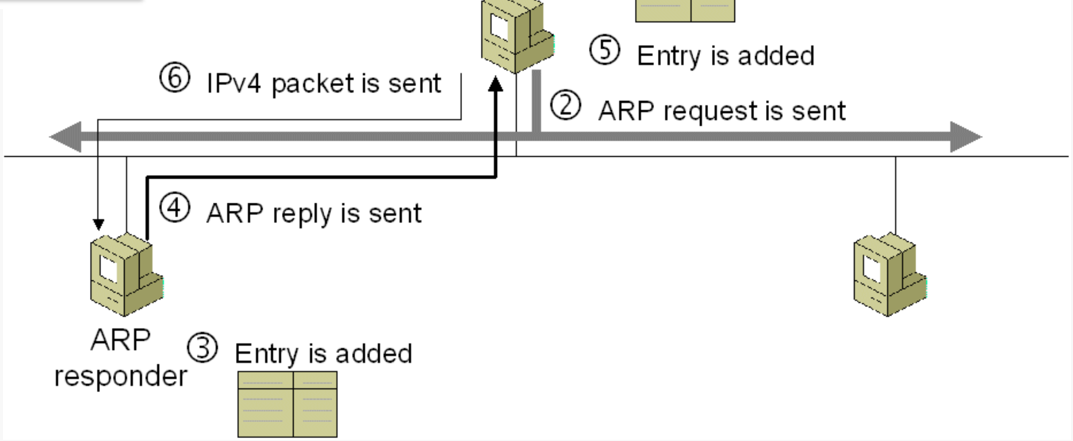
ARP header

0 7 15 31

| Hardware Type | | Protocol Type |
|---|---|---|
| Hardware Address Length | Protocol Address length | Opcode |
| Sender Hardware Address | | |
| | | Sender Protocol Address (bytes 1-2) |
| Sender Protocol Address (bytes 3-4) | | Target Hardware Address |
| Target Protocol Address | | |

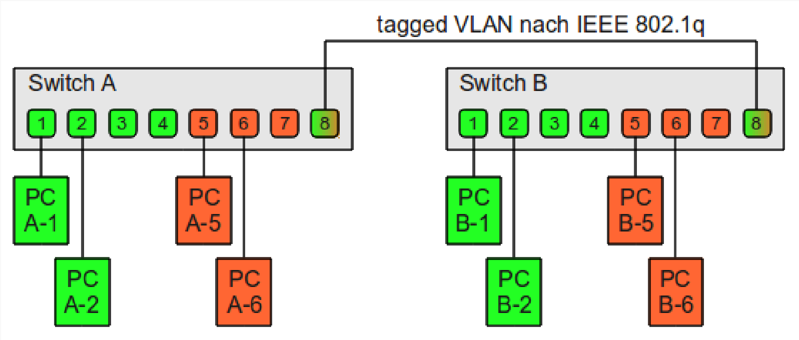


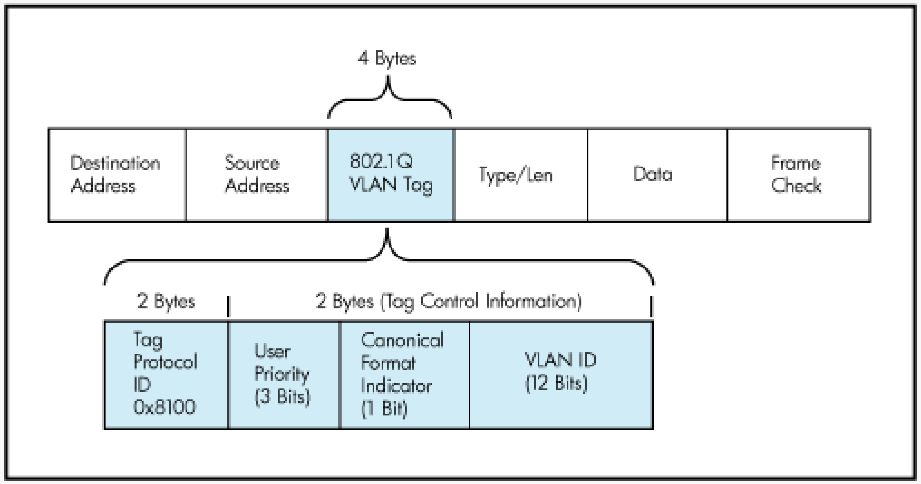
| TABELA ARP | |
|---|---|
| IP | MAC |
| 192.168.10.5 | 00:0A:B3:45:34:AB |
| 192.168.10.34 | 00:0A:B3:50:54:14 |
| 192.168.10.240 | 00:0E:F3:50:00:FA |



# Isolation: Virtual LAN (VLAN)

- Isolated virtual network segments: VLANs (IEEE 802.1Q)
  - without Layer3
  - better scalability





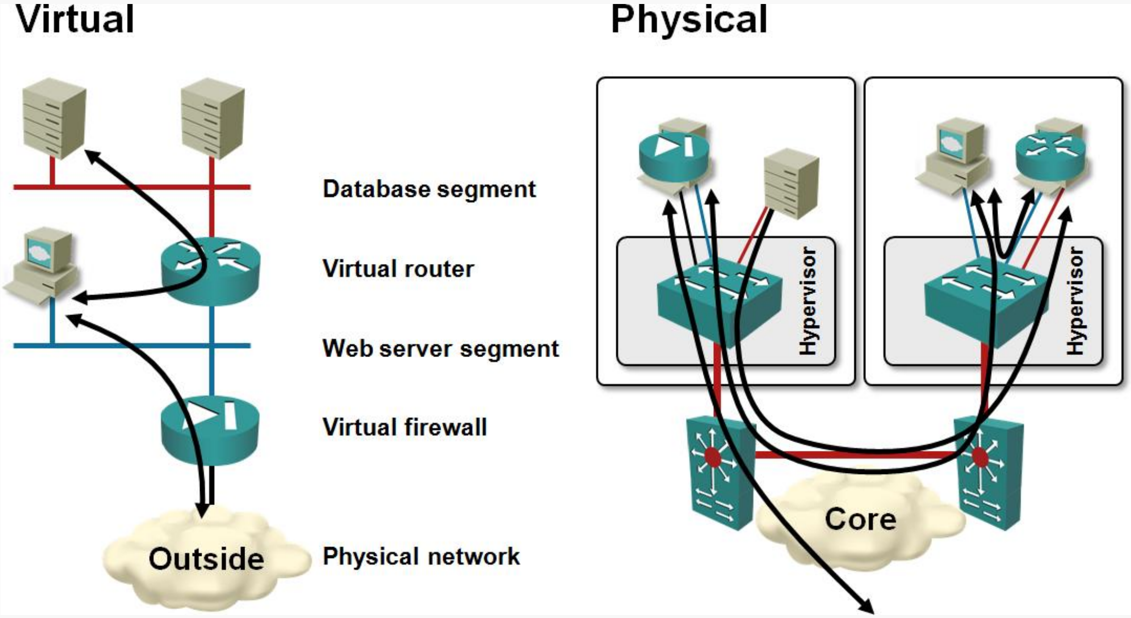
- Forwarding is based on VLAN ID and destination MAC
- Ethernet Network Interface Card (NIC)
  - MAC address filtering
    - for one or several unicast and multicast addresses, processes only frames destined here
  - Virtual Machines (VMs) on a physical machine (PM)
    - many VM (and corresponding MAC addresses) on the same PM
    - hypervisors usually set the physical NIC to „promiscuous mode” (accepting frames without filtering)
      - each frame is processed with the help of the CPU

# Scalability of VLANs

- Maximum 4094 VLANs in an Ethernet network
  - 12 bits VID (0x000 and 0xFFF reserved)
- hypervisor physical NIC in promiscuous mode
  - flooded frames are processed by CPU
- Usual implementation
  - all VLANs available on all server NICs
    - hypervisor processes all flooded frames, even if there is no active VM in that VLAN on the host
  - it is like we had only one VLAN

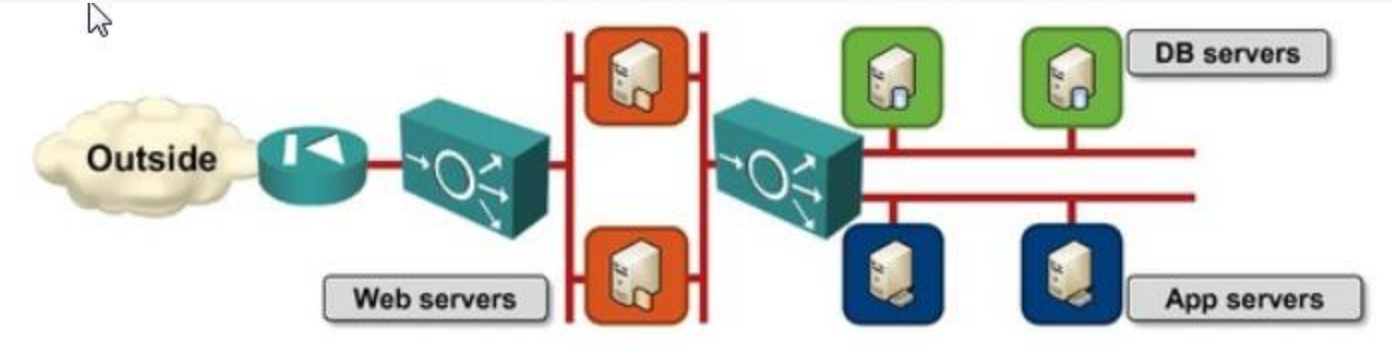
# Virtual Network Architectures

- One physical network – many virtual networks
- Many customer in the data center
  - each customer has many VMs
  - must work like they are on a private network
  - adaptation to the changing needs
- Tunneling, encapsulation
  - one or more tag



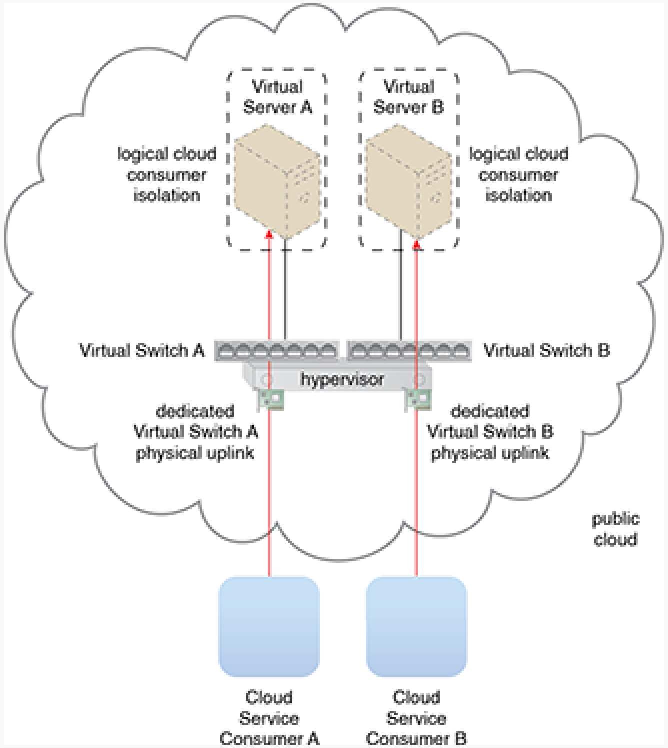
# Web Application Architecture

- A complex application requires network functions too
  - L2/L3 packet forwarding in multiple subnets
  - firewall
  - load balancing
  - NAT
  - VPN access



# Web Applications in the Cloud

- » For multiple customer all applications must be separated from the others
- » Keeping the existing network connections in operation
    - » internal addressing
    - » network services
    - » security modell
  - » virtual segments
  - » QoS

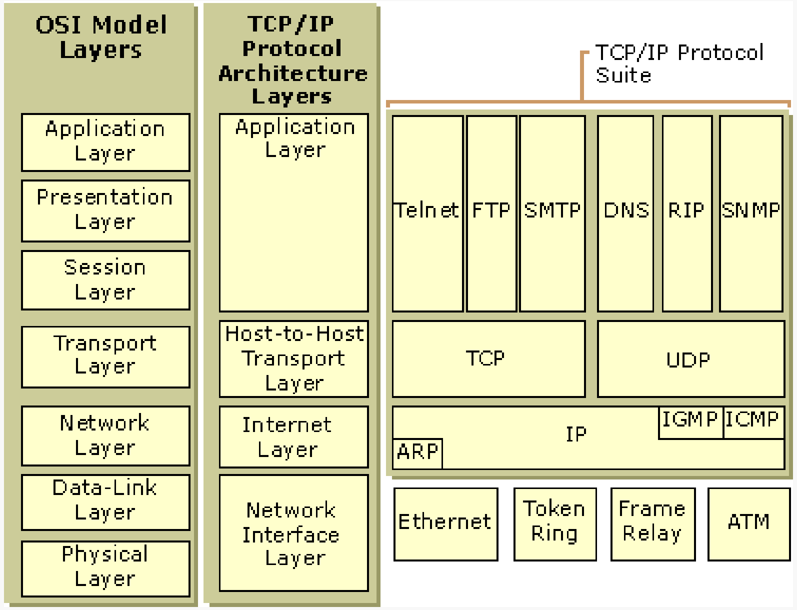

Virtual Server A
Virtual Server B
logical cloud consumer isolation
logical cloud consumer isolation
Virtual Switch A
Virtual Switch B
hypervisor
dedicated Virtual Switch A physical uplink
dedicated Virtual Switch B physical uplink
public cloud
Cloud Service Consumer A
Cloud Service Consumer B

# Virtual Network Architectures

- The important question: is it scalable?
- Goal: scalable infrastructure for several thousand virtual networks
- Scalability
  - keep the performance with the increasing workload by adjusting the processing capability
  - scaling up/down: vertical scaling
    - larger resource (faster or more CPU, larger or quicker memory and storage drive)
    - in the cloud the resources of the VM can be enlarged
  - scaling out/in: horizontal scaling
    - adding more server instances

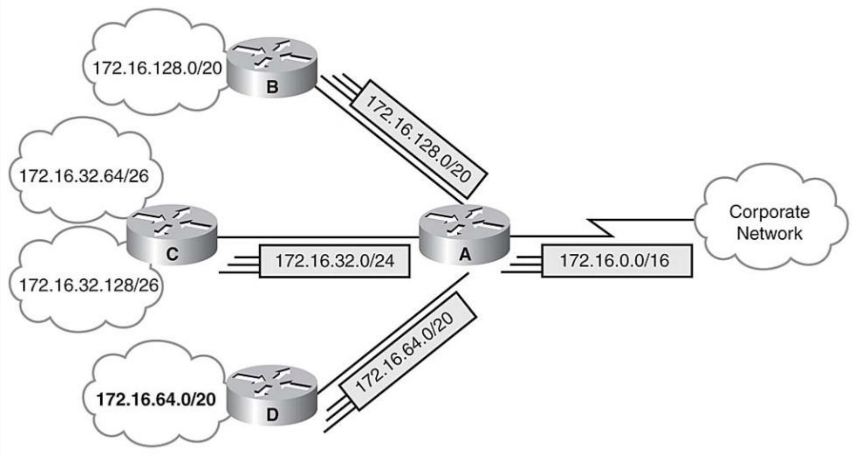
# Networking inside the Cloud

- Internet
  - world-wide, huge number of endpoints, it works quite good ☺
- Data center
  - similar requirements
  - even VMs in the order of million (e.g. AWS)
  - exponential growth
  - often the network bandwidth is the bottleneck
  - Options
    - Layer2
      - switching
      - simpler, plug-and-play
      - VM migration keeps the IP address
      - scalability?
        - up to small and medium size
      - typical in enterprise data centers
    - Layer3 (Amazon, Facebook, etc.)
      - routing
      - scales well
      - for any network size
      - however not a “small Internet”



# Networking in the Cloud

- Options
  - Layer2: Ethernet
    - MAC address is location independent
      - flat addressing
    - scalability limit: learning all the MAC addresses in the switches
  - Layer3
    - hierarchical address space
    - routing information is aggregated



# Networking in the Cloud

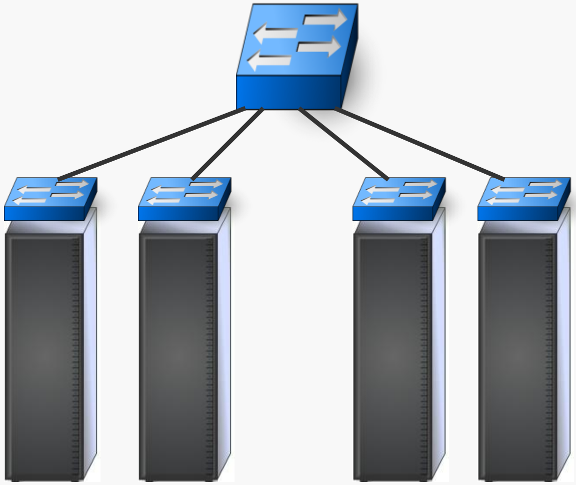
- Layer2: Ethernet
  - easy to configure and deploy: plug and play
  - approx. up to 1000 servers
  - communication within the local segment
    - traffic destined outside of the segment is sent to the default gateway
  - customer can manage the allocated IP address range
    - starting new VMs
    - change, reallocate  IP addresses
  - Spanning Tree Protocol
    - no multipath

# Networking in the Cloud

- Layer3
  - each network device is a router
    - protocol: Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS)
    - distributing topology information
  - one VM – one L2 “network”
    - no L2 broadcast, multicast is difficult
    - no VLAN
    - e.g. Windows servers use broadcast to discover each other
  - Equal Cost MultiPath (ECMP)
    - better network bandwidth utilization
  - shortest path
    - Dijkstra’s algorithm
  - VM migration is more complex
    - IP address change

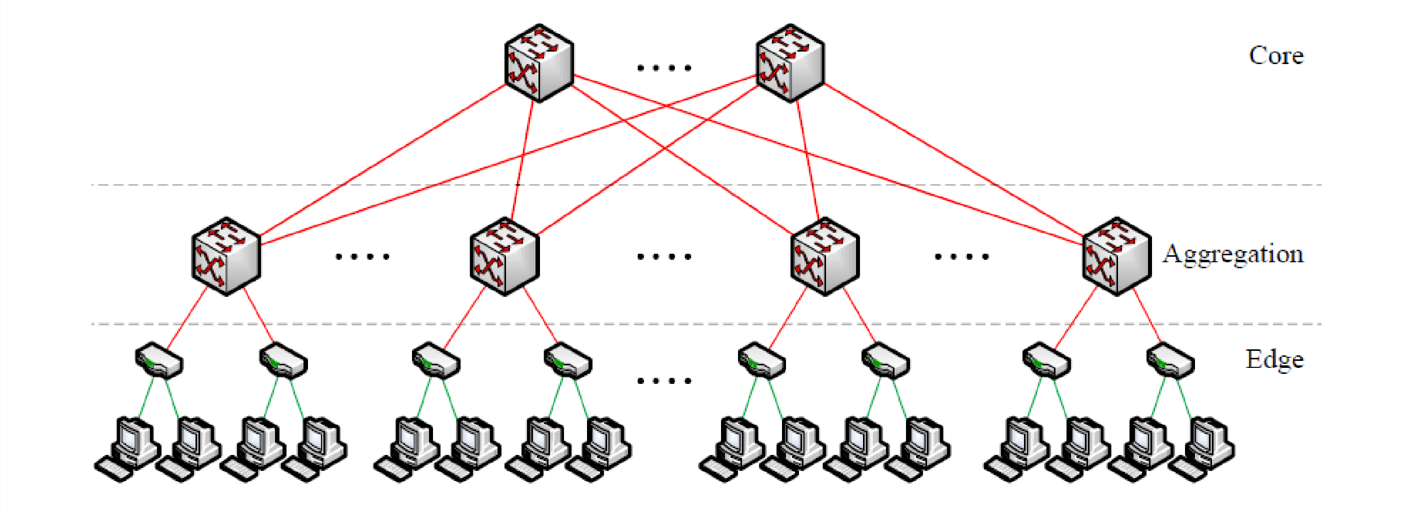
# Network Topologies

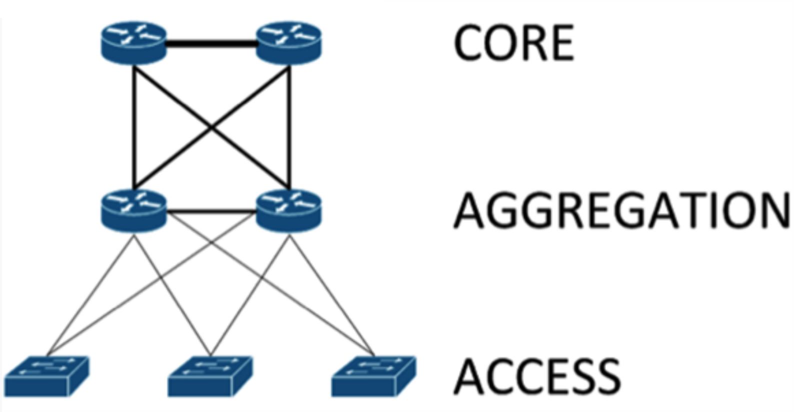
- 3 level hierarchy: ToR, aggregation, core switch
- flat(ter) topology, 2 levels: ToR and core switch
  - one central switch: expensive, port number is limited
    - e.g. the price of a 128 port GbE switch is approx. 100 times the price of a 48 port switch



# Network Topologies

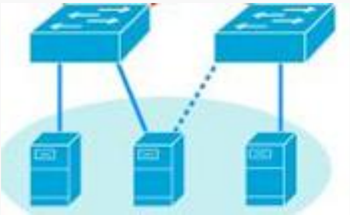
- Redundancy and/or load balancing
  - dual star

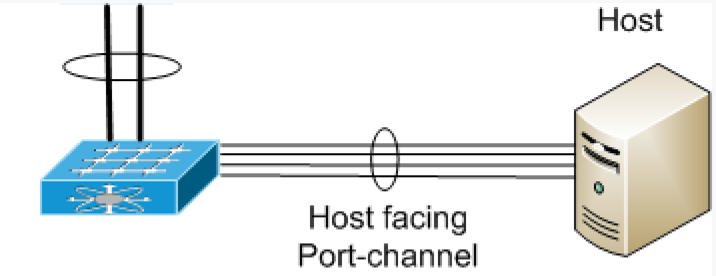




# Server – ToR Switch

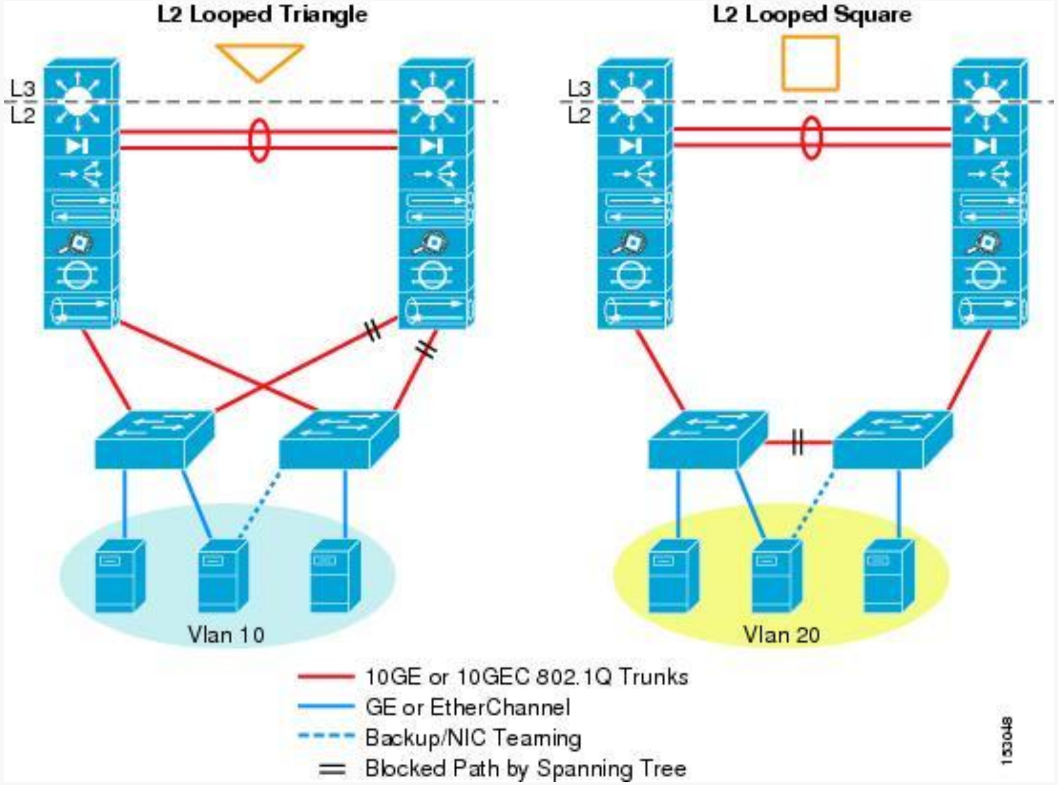
- Single homed
  - 1 server – 1 switch
  - resiliency can be provided by external mechanisms
  - Single Point of Failure
    - NIC, cable, switch port, switch
- Multi homed
  - 1 server – 2 switches
    - working/backup, primary/secondary
    - simultaneous operation
      - different MAC, IP addresses
- Port-channel
  - requires switch configuration
  - one logical connection, aggregation of more physical links
  - 1 server – 1 switch
  - 1 server – more switches
    - virtual port-channel
    - ending on different physical switches
    - shared or communicating control planes
    - switches are also connected
      - forming a ring if there are more than 2 switches





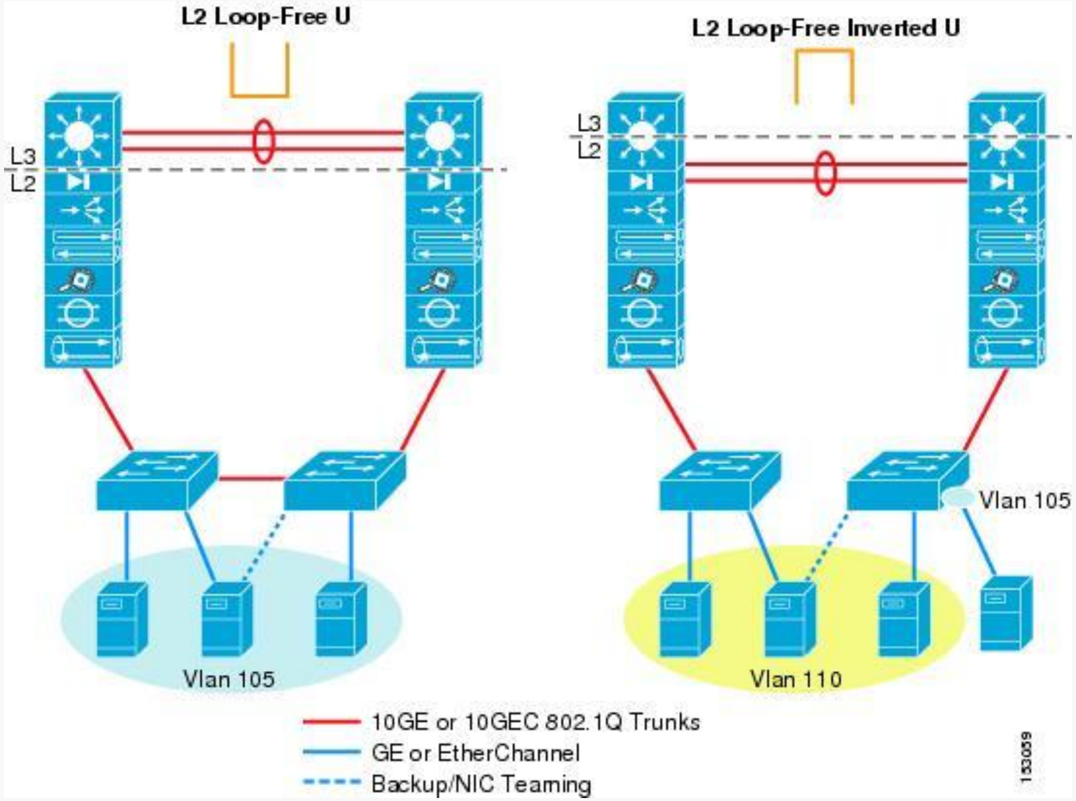
# ToR – Core Switch

- Looped Topology
  - triangle
    - wide-spread solution
    - half of the connections are unused
    - more ports on the core switch
  - square
    - less redundant
    - less ports on the core switch



# ToR – Core Switch

- Loop-Free Topology
  - STP not running
  - U
    - ToR switch operates in transit mode in case of failure
  - inverted U
    - less ToR switch port
    - not applicable for single homed servers
    - without network level redundancy


L2 Loop-Free U
L2 Loop-Free Inverted U
L3
L2
L3
L2
Vlan 105
Vlan 110
Vlan 105
10GE or 10GEC 802.1Q Trunks
GE or EtherChannel
Backup/NIC Teaming
153039