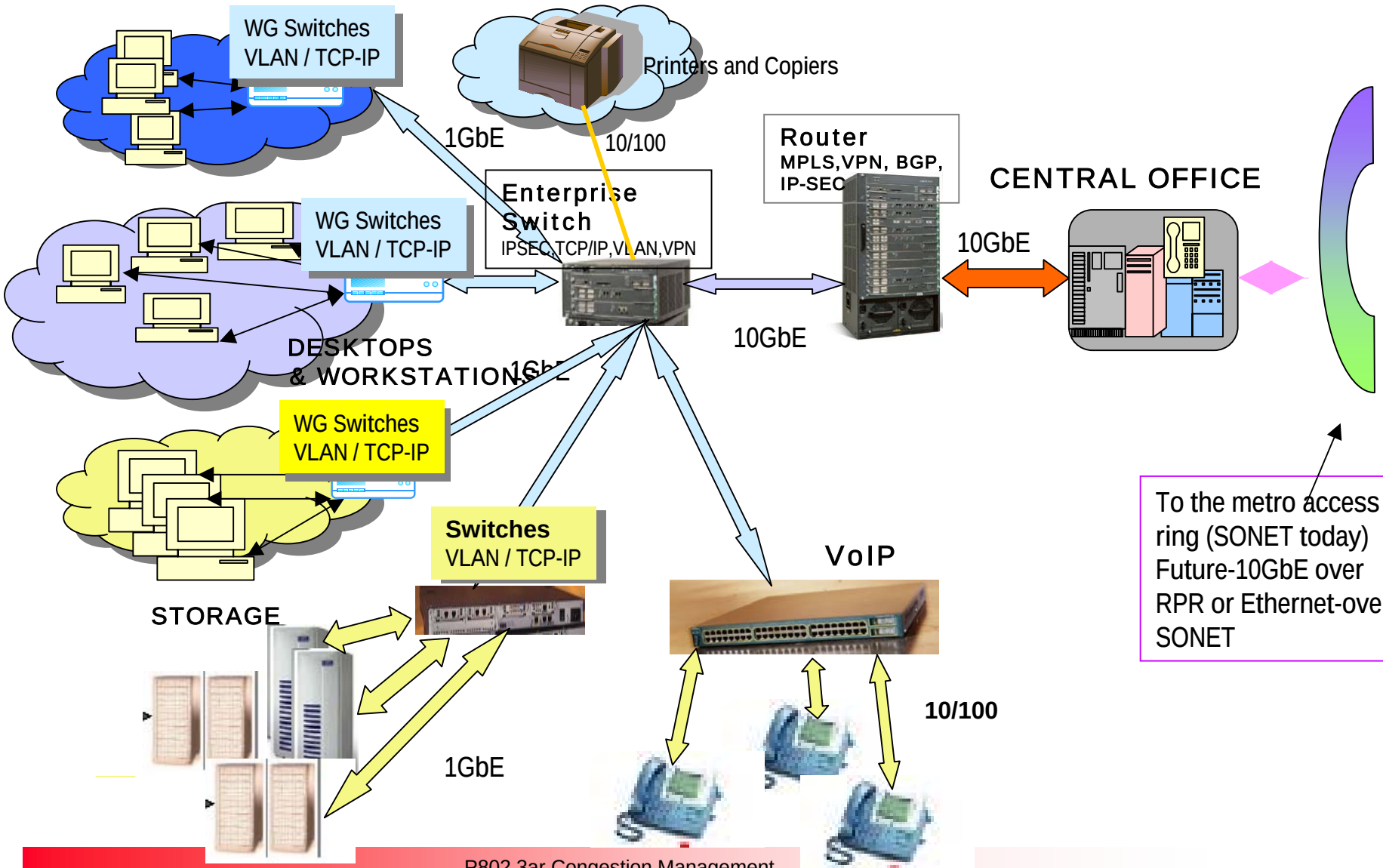
# Congestion Management in Backplanes and the Enterprise

**Asif Hazarika**

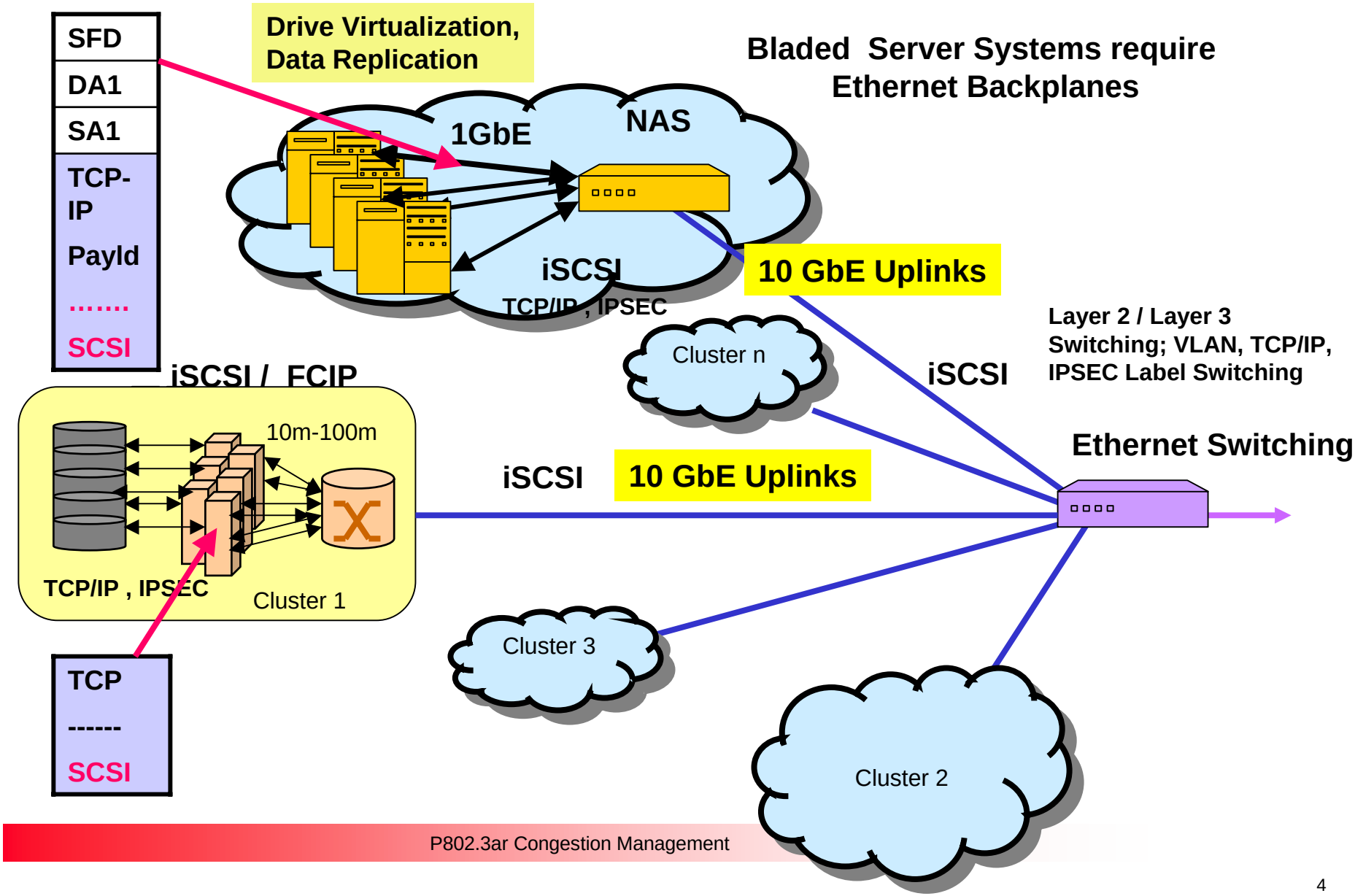
**ahazarik@fma.fujitsu.com**

# Today’s enterprise network

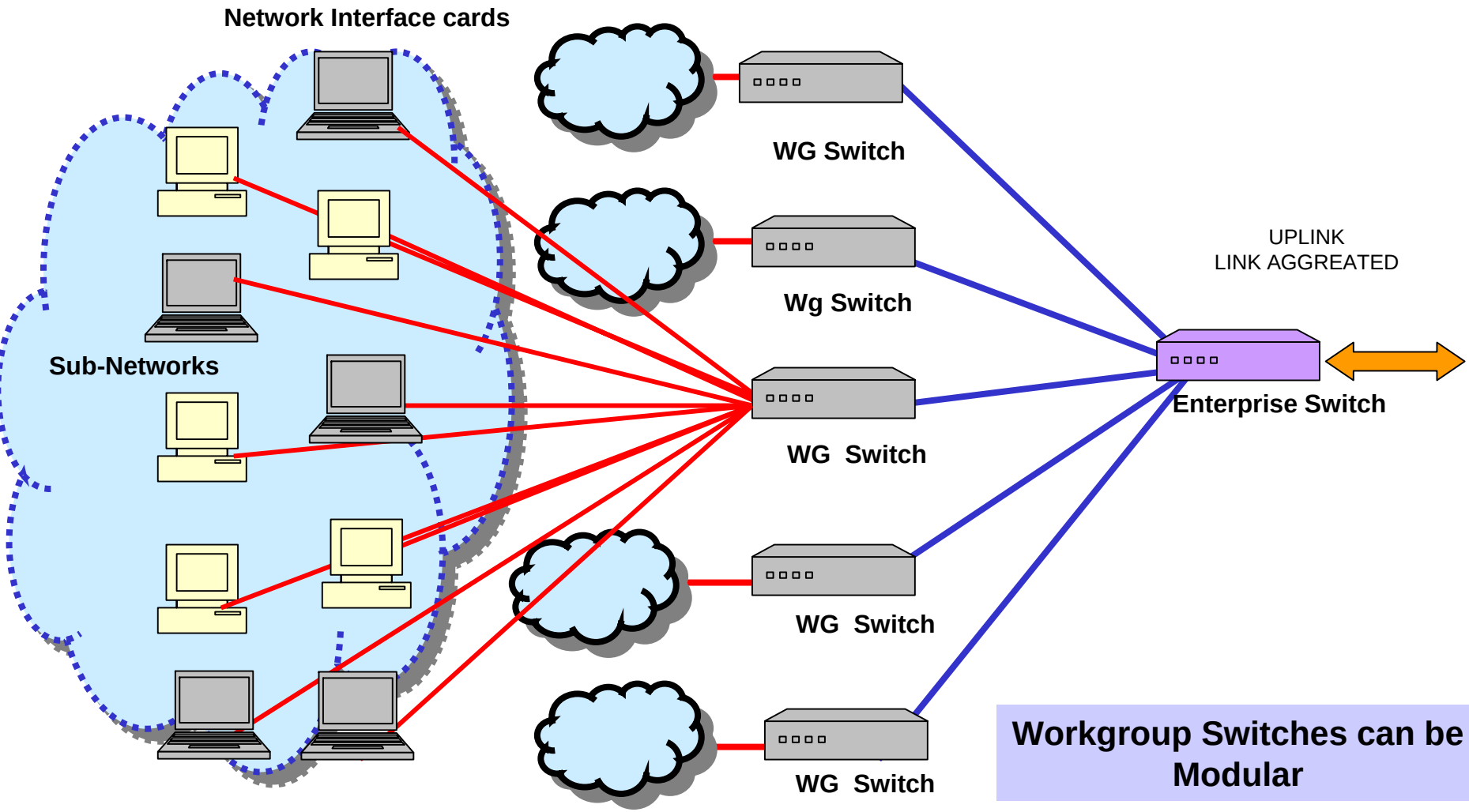
# The Enterprise Network



# Storage area networks



# Enterprise networks



# Switching &other Networking Equipment - Backplanes

- **High Performance Equipment utilize a backplane infrastructure – to allow flexibility and scalability**
- **Modular Switches**
  - 1Gbe Switches with 96-port , 144-port, 288-port and beyond need high performance Ethernet backplanes
- **Bladed Servers**
  - Using Ethernet Backplanes – today at 1GbE X 4
- **ATCA Chassis**
  - Are 1GbE backplanes going on to 10GbE

# Network Topology and 802.3ar

- **What is the scope 802.3ar ?**
  - PAR:
    - Specify a mechanism to support the communication of congestion information
    - Specify a mechanism to limit the rate of transmitted data on an Ethernet link
    - Preserve the MAC/PLS service interfaces
    - Minimize throughput reduction in non-congested flows
  - Which network segments should ‘ar’ address?
    - Although there is end-to-end Ethernet possible should we address the solution across all these domains? - No
    - Narrow down to a Client Server model or a Peer to Peer model

      and Backplanes where Ethernet is used

# What causes congestion?

- **Congestion**
  - Happens at Aggregation and Switching points
  - Due to over-subscription of capacity of a network elements or node
  - Link failure causing reduced availability of downstream bandwidth
- **Why over-subscription?**
  - Designer take advantage of Bursti-ness of the data
    - Count on the fact that the load factor will never achieve 100%
    - Advantage : Lower system costs (Use traffic management)
    - Disadvantage : Period of coincidental Peak Bandwidth cause lossyness

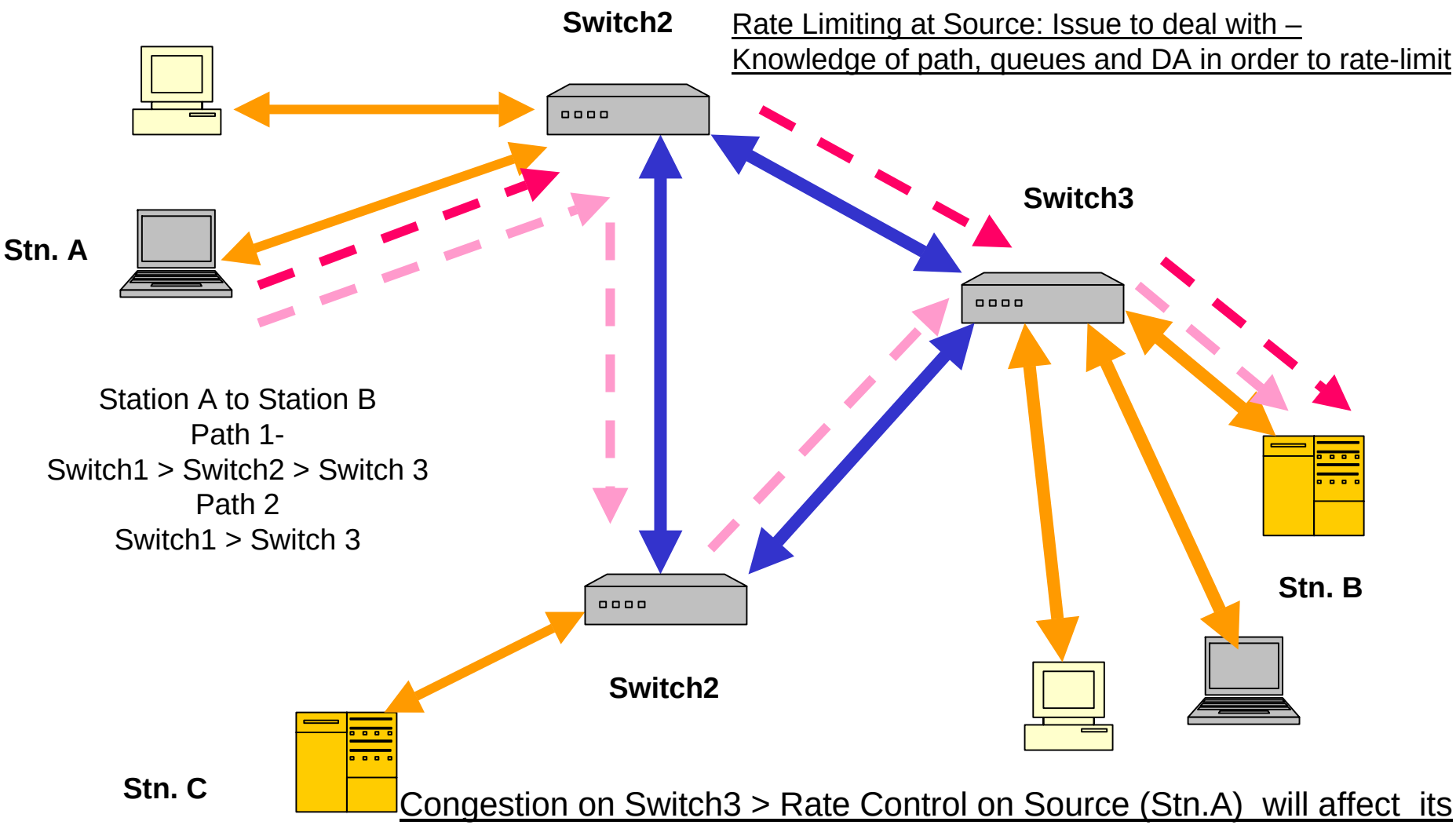
# Congestion Management

- **Congestion Management**
  - Congestion Avoidance
    - Over-provisioning
    - Predicting congestion and taking steps before congestion happens
  - Adaptive Queuing Mechanism
    - Discard packets before threshold are reached
    - Signal to TCP-IP windowing mechanisms to reduce traffic
  - Credit based system
    - Mechanism to issue credits to Fabric Interface devices to control rates at the output queues
    - Messaging between Fabric and the upstream scheduling point.
  - Peer to Peer or Client Server application can be reasonably managed in today's networks…. However in a backplane environment is where we see real issues with Ethernet

# Current proposals

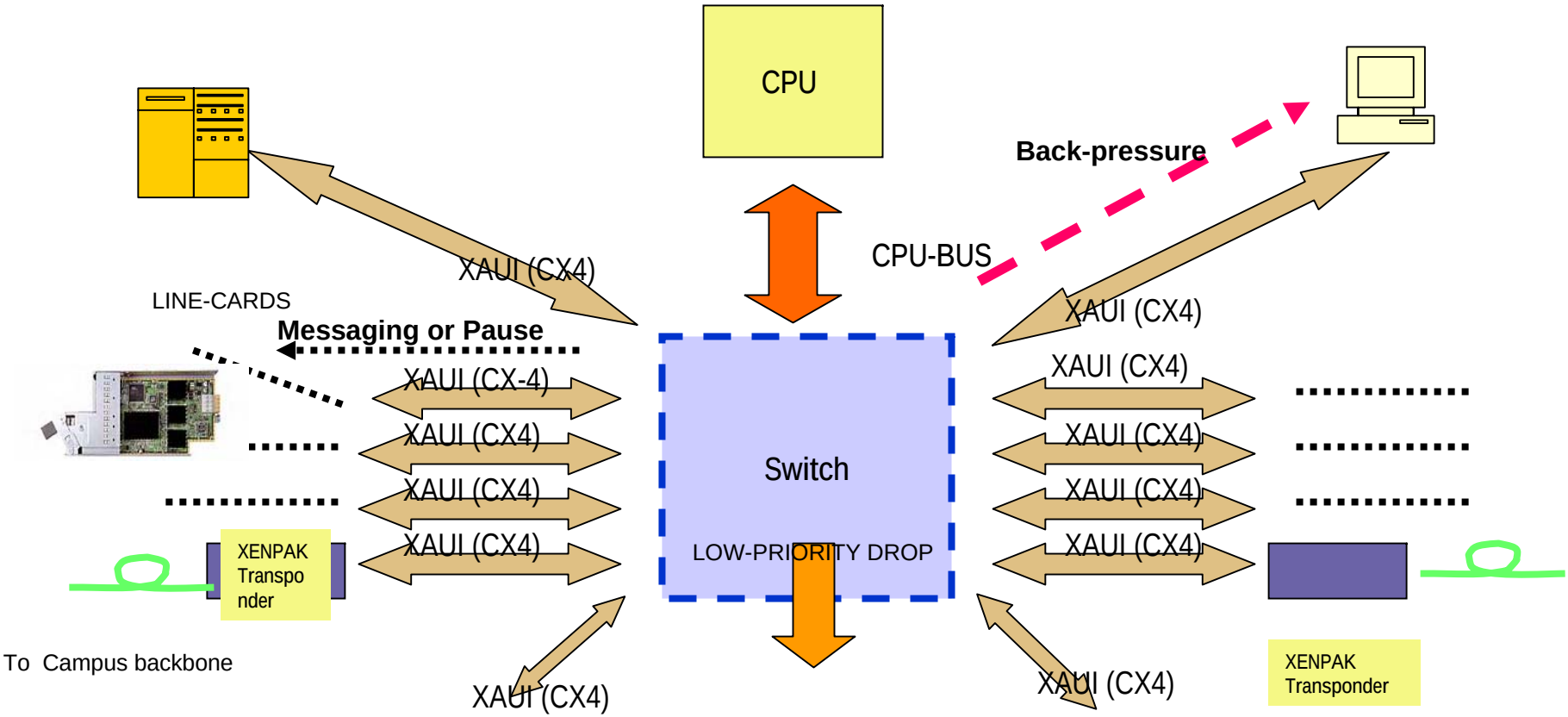
- **Using a Congestion Messaging scheme in Enterprise**
  - Forward Notification
    - Intent : To have the end node transmit message to source node that congestion was encountered , and to limit the rate.
    - Advantage: Simple. BUT
    - Delay from the time Congestion  is experienced to when the source node gets the message.
    - Too much S/W interaction required
    - If 90% Traffic is TCP-IP will the other mechanisms work better?
  - Backward Notification
    - Too much traffic in a loaded environment
  - Rate Limiting: Multiple paths can be an issue – Many possible path, taken by a traffic if you rate control the end node you may do more harm then good
  - Rate Limiting: Important where it is done. It has to be done at the scheduler level
  - Rate control is good in a Frame relay environment or where a connection path is established

# Multiple paths



# Switch in various interconnect scenarios

Switch is the heart of any LAN or Chassis System and therefore a congestion point!




**Switch should have the capability to recognize congestion and :**

**1- Apply backpressure when congested**

**2- Drop low-priority traffic when congestion is experienced**

**3- Drop LP traffic when congestion is anticipated at certain buffer thresholds**

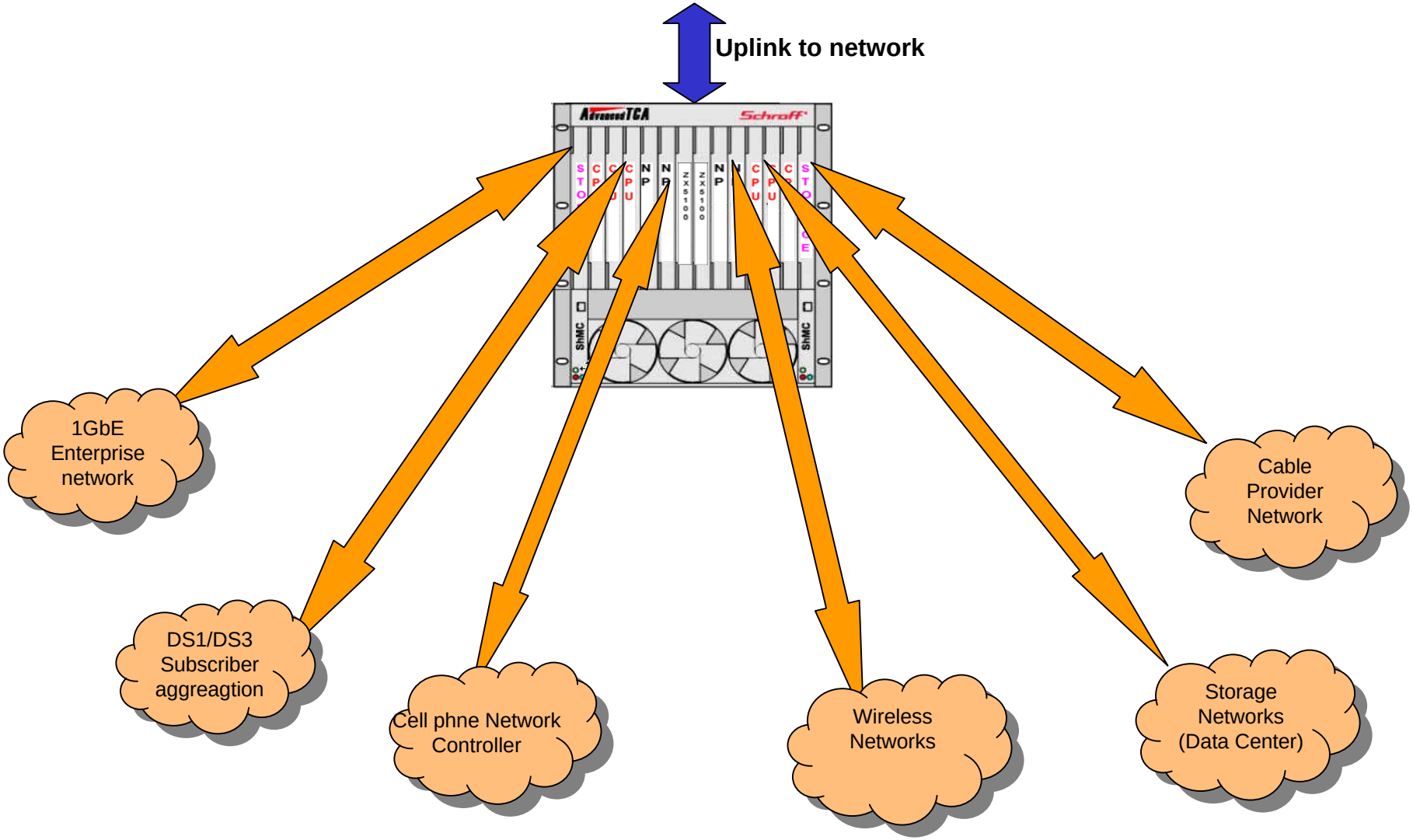
**4- Ability to force to narrow Windowing by TCP-IP**

**5- Send out , Congestion notifications/Control messages indicating congestion**

# 802.1p priority values

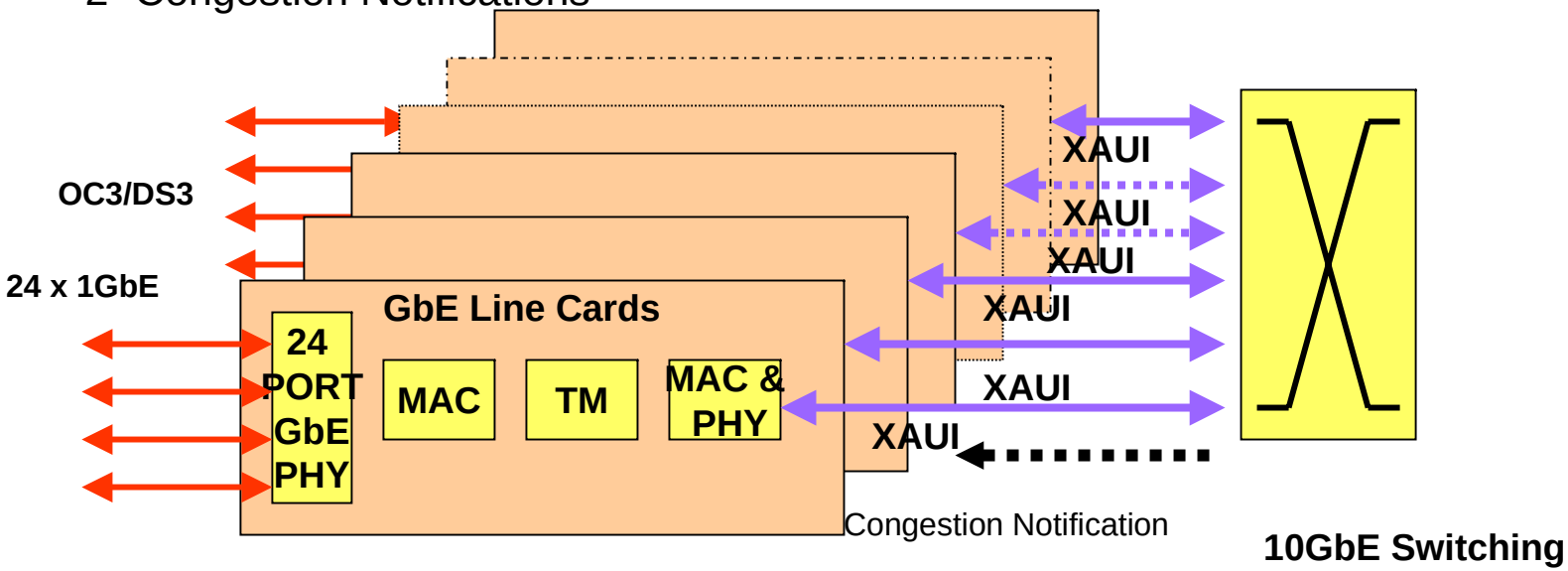
- **Mapping of Intserv to 802.1p (TCI field of the VLAN TAG)**
  - ‘7’ Network Control trafiic
  - ‘6’ Delay Sensitive  10-ms Bound
  - ‘5’ Delay Sensitive 100-ms Bound
  - ‘4’ Delay Sensitive no-Bound
  - ‘3’, ‘2’  Reserved
  - ‘1’  Reserved , Less than Best Effort
  - ‘0’ Default, assumed BE
- **Integrated Services consists of a packet scheduler, admission control, classifier and setup mechanism – reservation  setup  protocol**

# Network convergence



# Backplanes

Backplanes are unique:
It is confined to a physical location
Therefore control of end-points can be done through :
1- Network Management layer
2- Congestion Notifications




Congestion Notifications from the Switch will require the use of Pause and special Ether-types used for messaging:
1- To provide notification that issue credits to the schedulers
2- To provide messaging when congestion occurs.

# Conclusions

- **There are two areas of interests in Congestion Management**
  - LAN network (CM should be between two queuing points)
  - Backplane (CM between switch fabric and the line-cards)
- **Cannot expect to have a completely loss-less network**
  - Higher priority traffic could be loss-less
  - Lower priority traffic can be dropped, and the traffic in most cases , will be retransmitted (If TCP/IP)
- **Congestion avoidance can be realized in a backplane by managing the line-card output queues**
  - Congestion Notification to the output queues
- **Early discards using 802.1p priority tags to provide feedback mechanism to TCP/IP source can work in enterprise where there is TCP traffic**
- **We must have a mechanism in (802.1/802.3) to create Congestion Notification frame.**

# Thank you!