



# 802.17 presentation

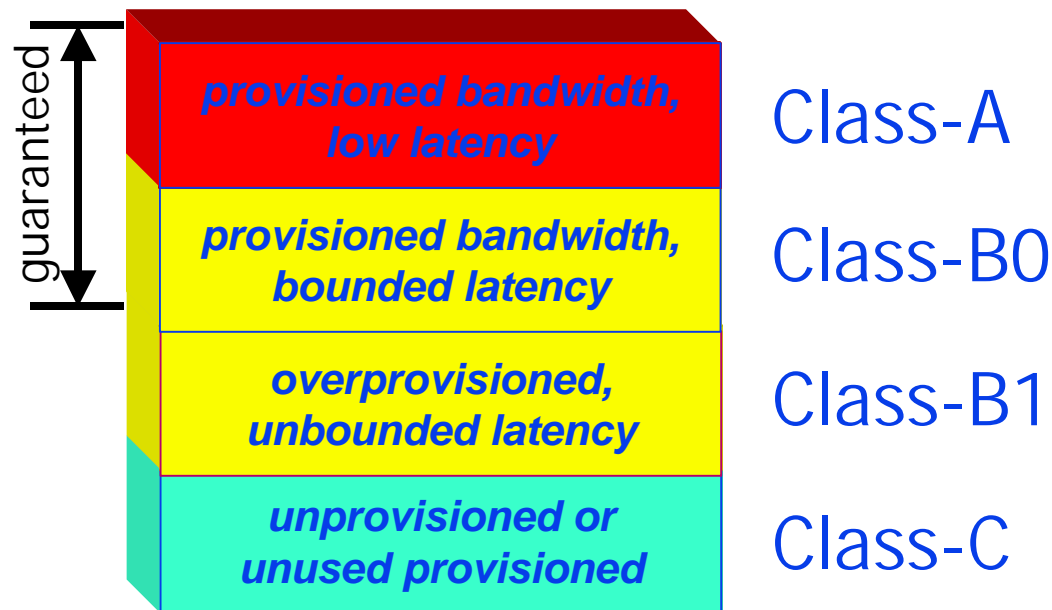
- Prepared for 802.17, November 2001
- Dr. David V. James  
Chief Architect  
Network Processing Solutions  
Data Communications Division  
110 Nortech Parkway  
San Jose, CA 95134-2307  
Tel: +1.408.942.2010  
Fax: +1.408.942.2099  
Base: [dvj@alum.mit.edu](mailto:dvj@alum.mit.edu)  
Work: [djz@cypress.com](mailto:djz@cypress.com)



# Flow control

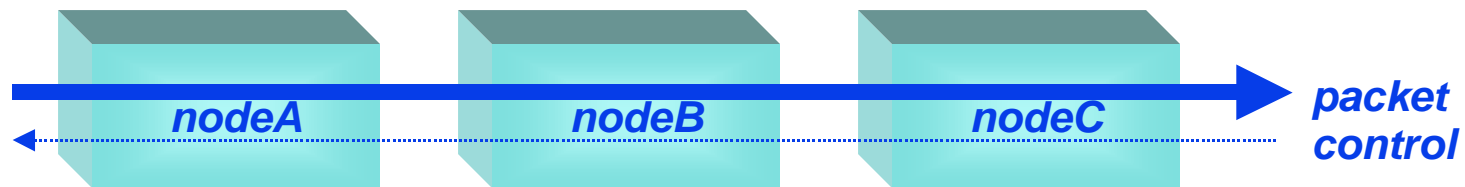


# Arbitration classes





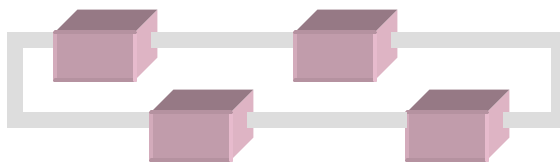
# Opposing arbitration



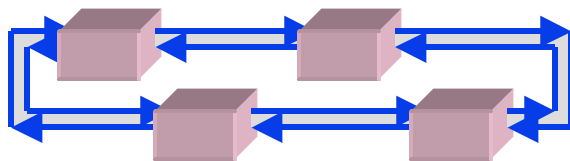
- Data packets flow in one direction
- Arbitration control flows in the other\*



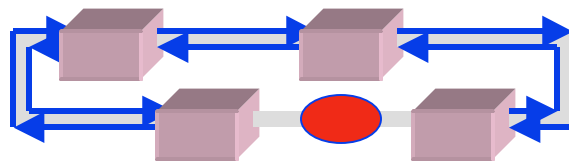
# Supported topologies



- A physical ring



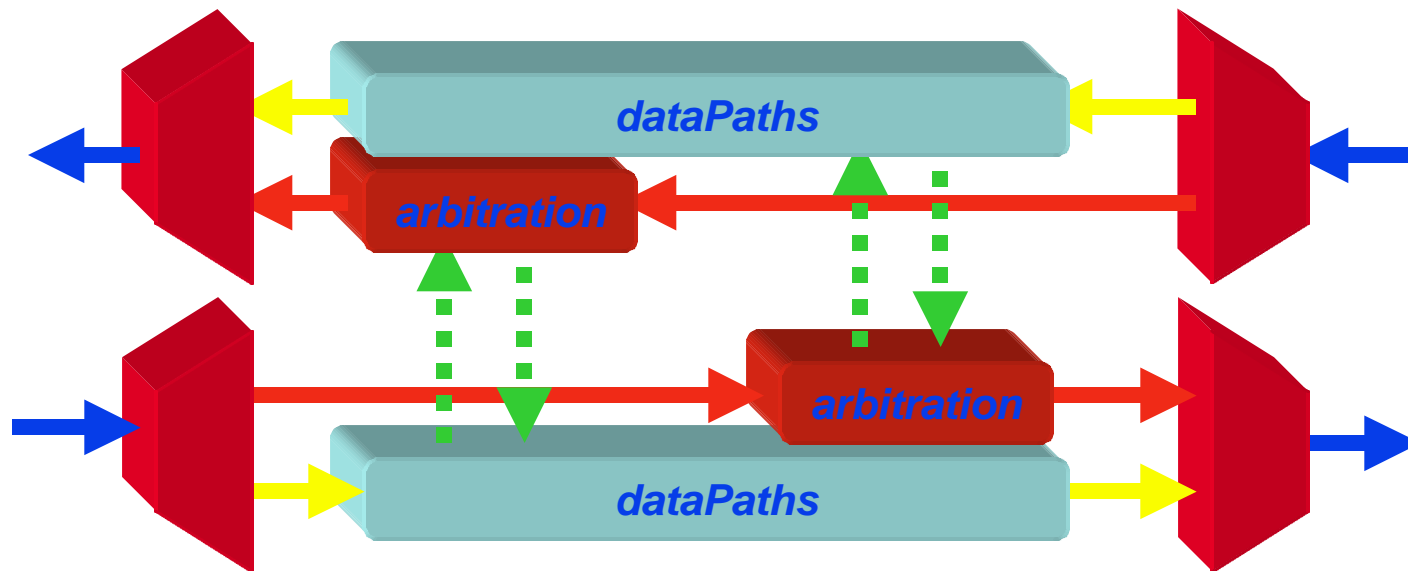
- Dual ringlets



- Duplex ringlet



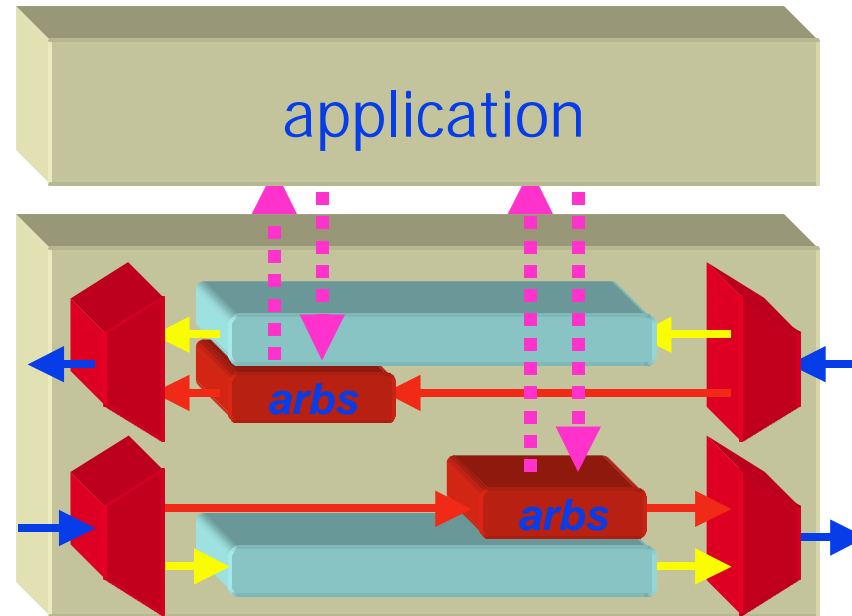
# Internal MAC arbitration signals



- Arbitration affects opposing run
- My congestion affects upstream node
- Downstream congestion affects me



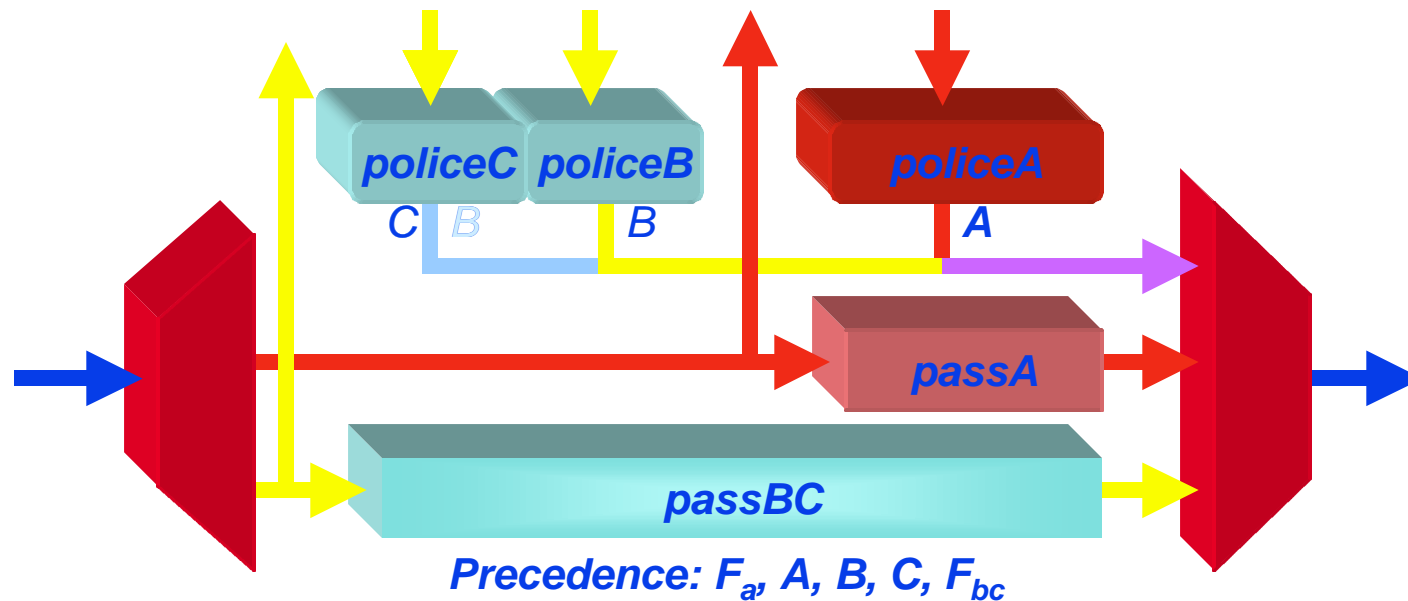
# External MAC arbitration signals



- MAC receives information
  - MAC FIFOs are \$\$, latency++ , inflexible
- Application receives information
  - Allows reordering and run selection



# Arbitration related components



- Distinct class-A & class-B/C paths
- Load dependent policing





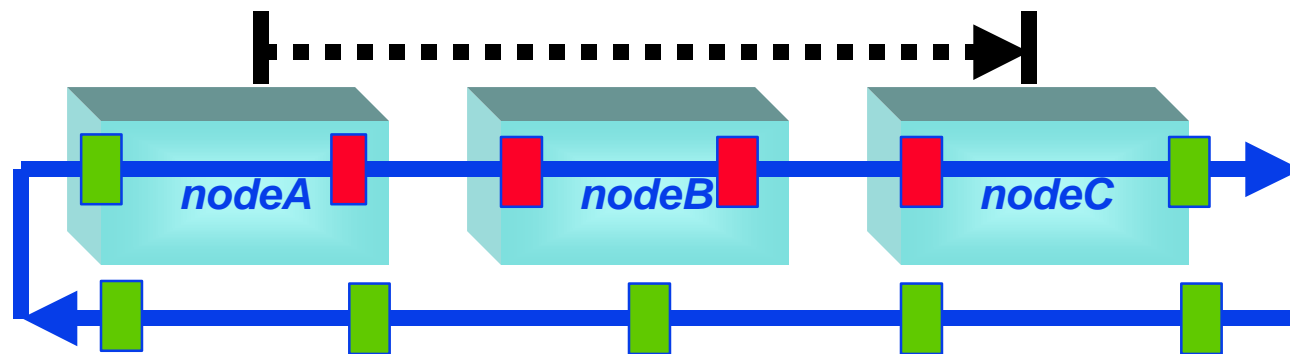
# **Class-A flow control (proactive and reactive)**

# Class-A flow control

- **Proactive**  
Minimal (nonexistent?) passBC transit buffer  
Less available bandwidth  
Each station maintains constant classAp traffic
- **Reactive**  
Significant passBC transit buffer  
Full bandwidth utilization  
Each station responds/regenerates throttle messages
- **Interoperable?**  
This is a bandwidth vs memory \$\$ tradeoff



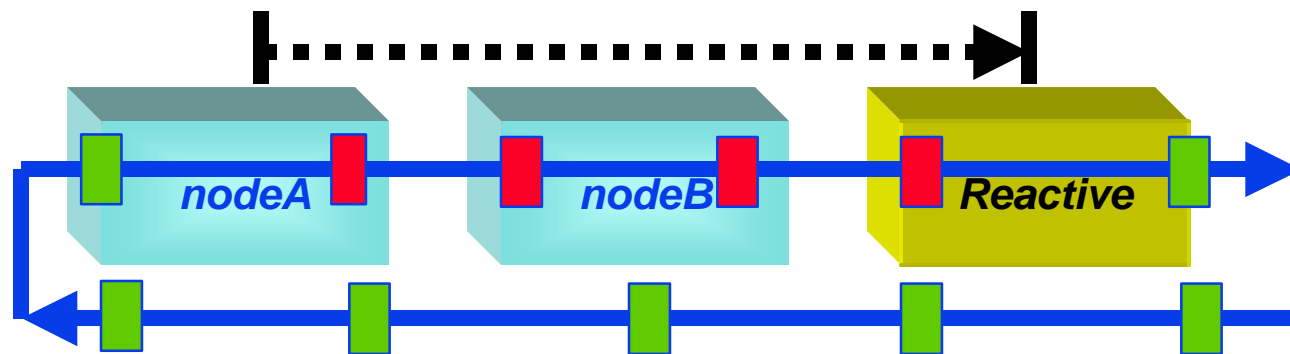
## Proactive class-A partitions



- Data packets go source-to-destination
- Residue returns destination-to-source to provide subsistence for transmissions



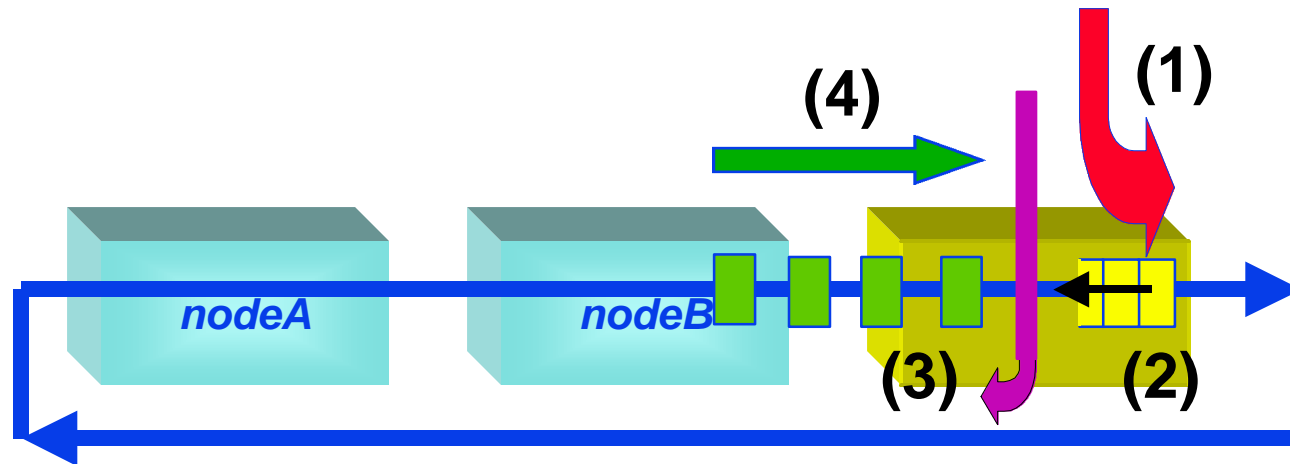
# Proactive class-A compatibility options



- Reactive node trickles class-A bandwidth
- Reactive node recycles class-A bandwidth  
class-A => class-A', thus preserving BW



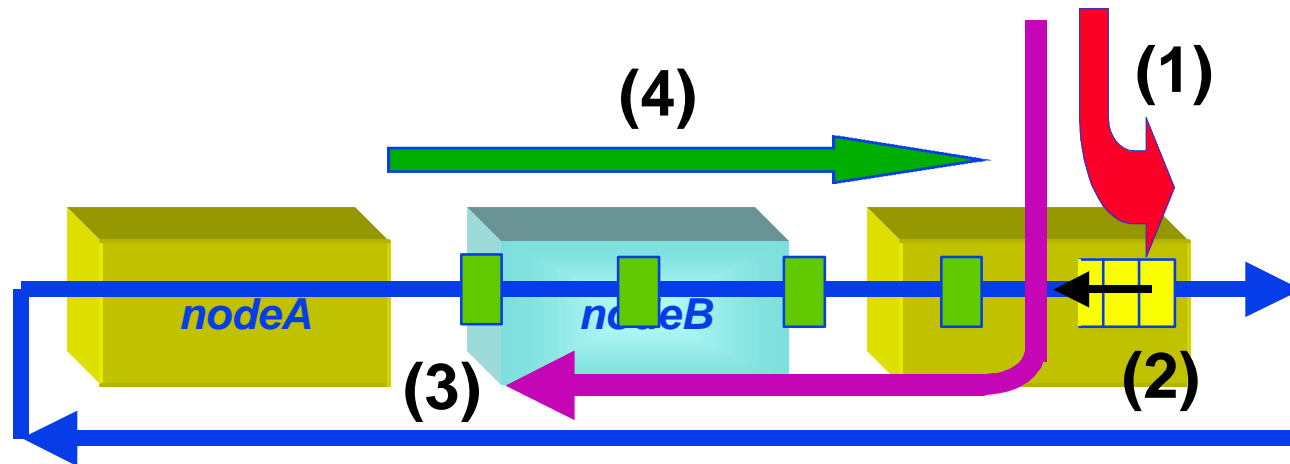
## Reactive class-A control



- Transmission of packets causes
- Backup of passBC FIFO that
- Returns flow-control information that
- Provides consumable idle packets



# Reactive class-A compatibility



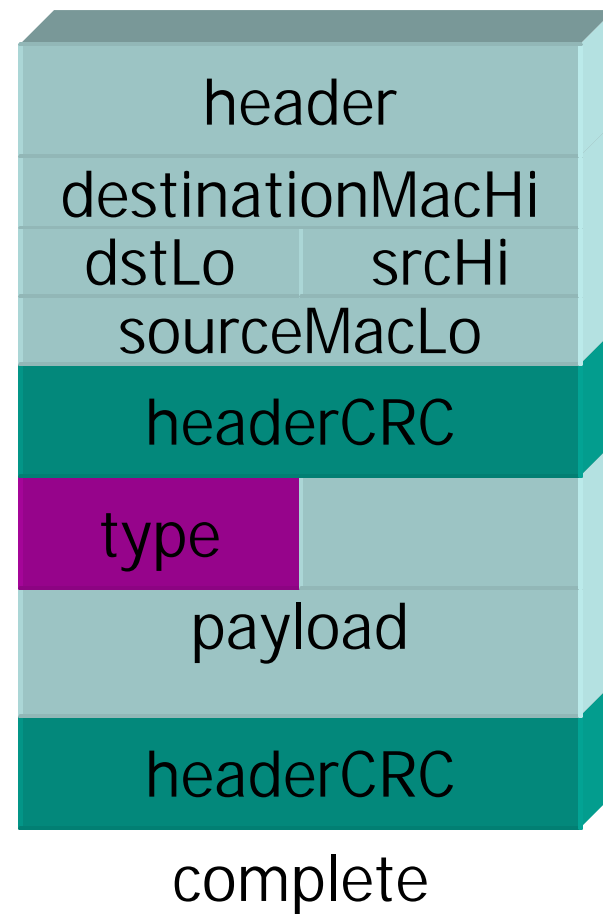
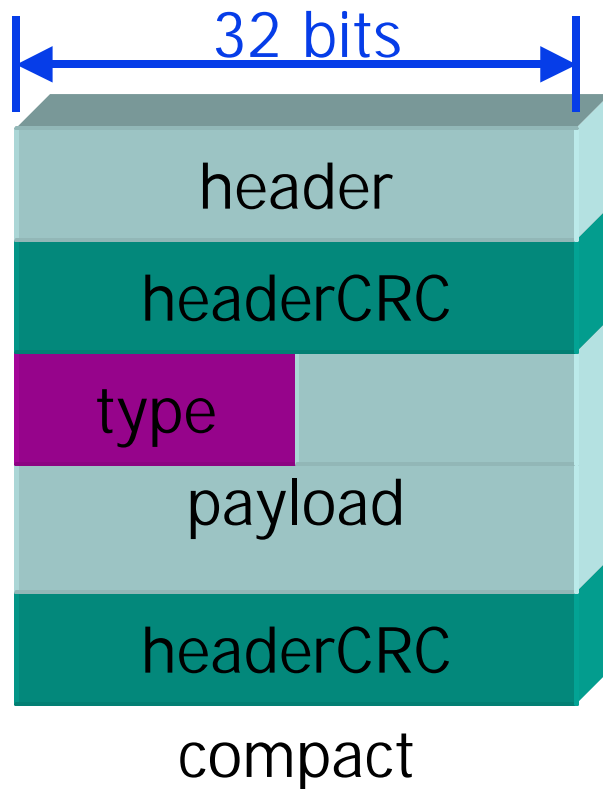
- Flow control passes upstream
- Proactive stations pass these indications



# Frame formats

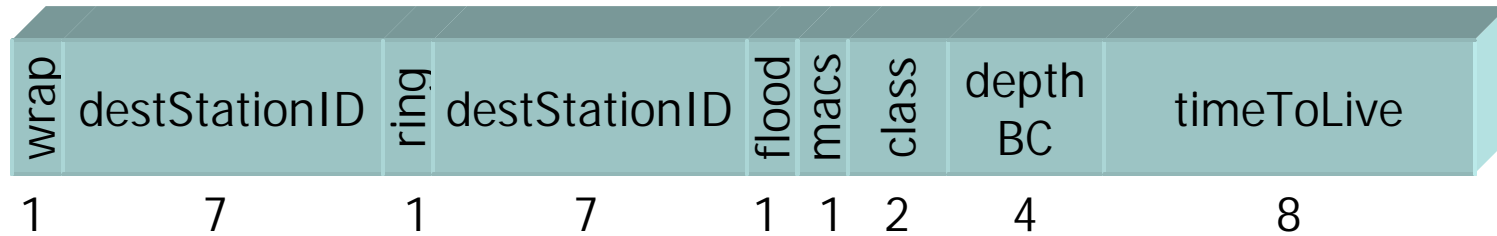


# Frame formats



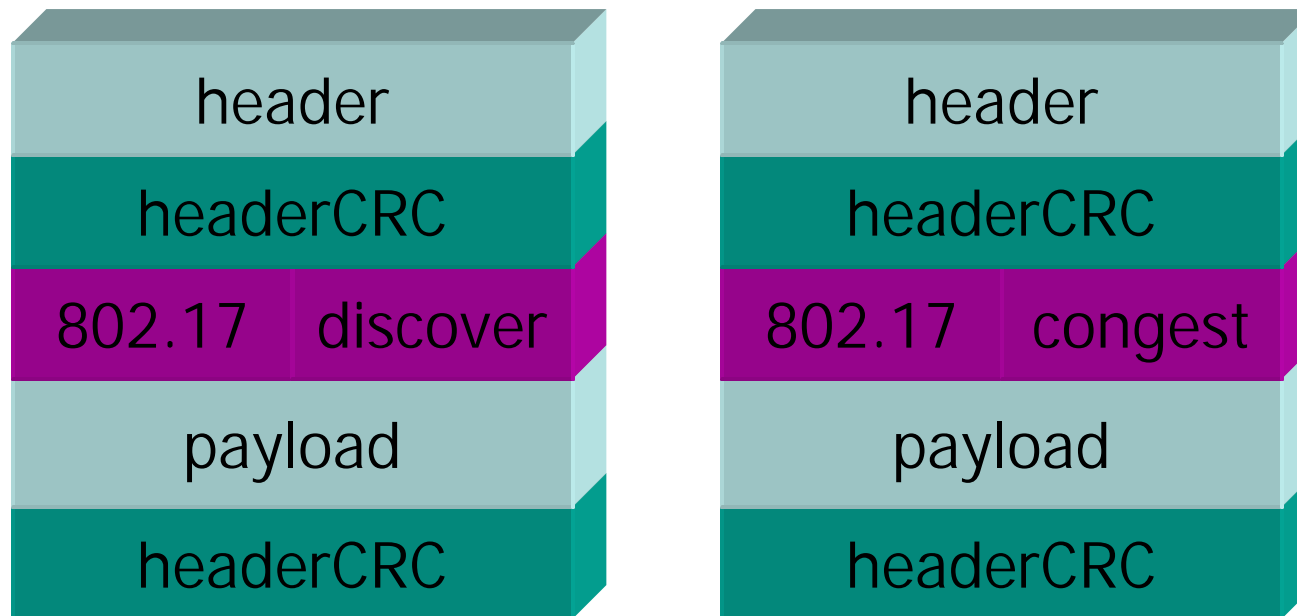


# Header format





## RPR control frames





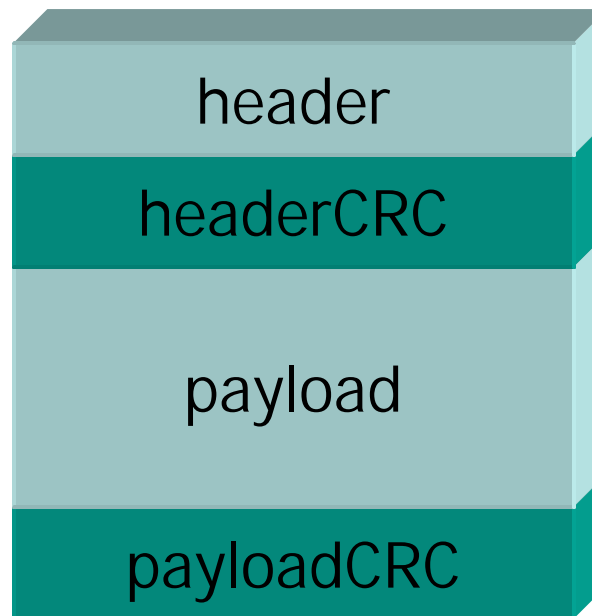
# CRC processing



# CRC processing

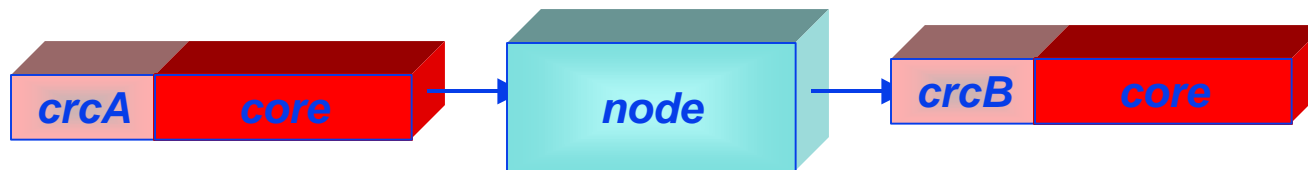
- **Store&forward/Cut-through agnostic**
- **Invalid data is effectively discarded**
  - **store-and-forward discards**
  - **cut-through stomps the CRC**
- **Maximize error-logging accuracy**
  - **Separate header&data CRCs**
  - **“most” corruptions hit the data**

# Separate header and data CRCs





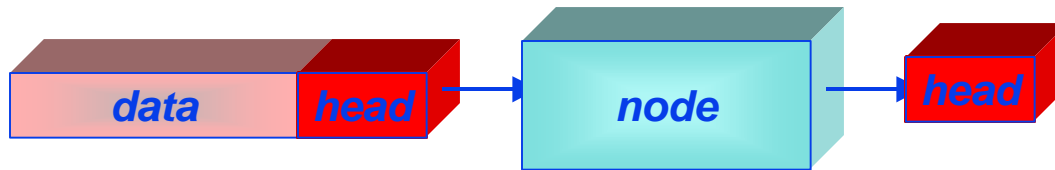
## Cut-through CRCs



- Corrupted packet remains corrupted
- Error logged when first detected
- ```
if (crcA!=crc) {  
    errorCount+= (crcA!=crc^STOMP);  
    crcB= crc^STOMP;  
}
```

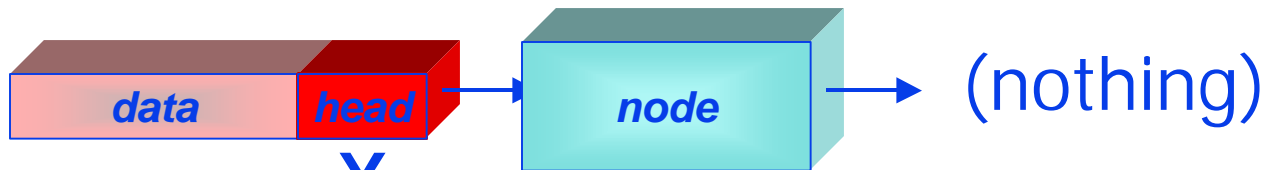


## Distinct CRCs reduces discards



X

- Discard the corrupted data

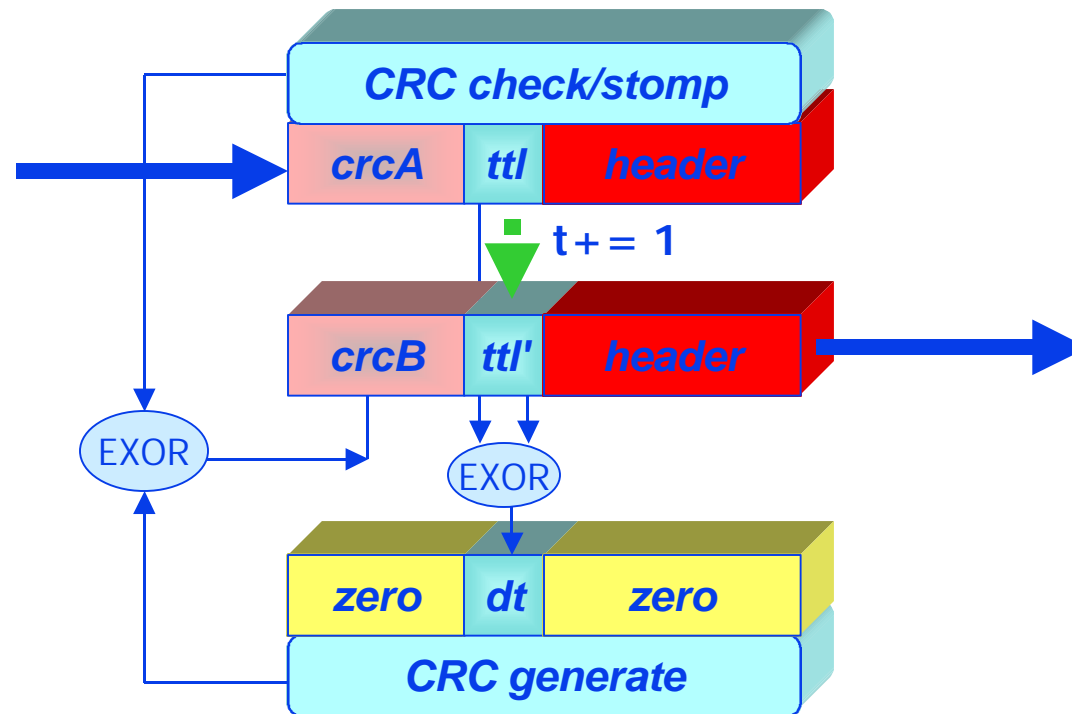


X

- Discard the corrupted packet



# End-to-end CRC protected TTL







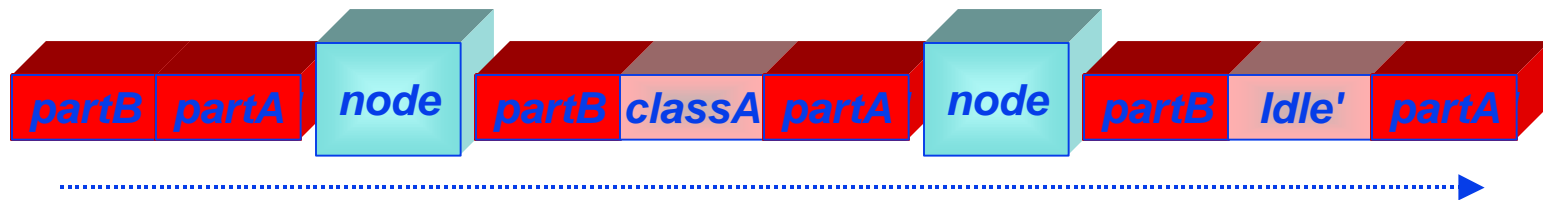
# **Pre-emption (a physical layer decision)**

# Pre-emption

- **Suspend class-B/C for class-A packet**
- **Only one level is sufficient**
  - **class-A is the latency critical traffic**
  - **more levels complicate hardware**
- **Physical layer dependent**
  - **marginal for high BW & small packets**
  - **distinctive “suspend” symbol required**



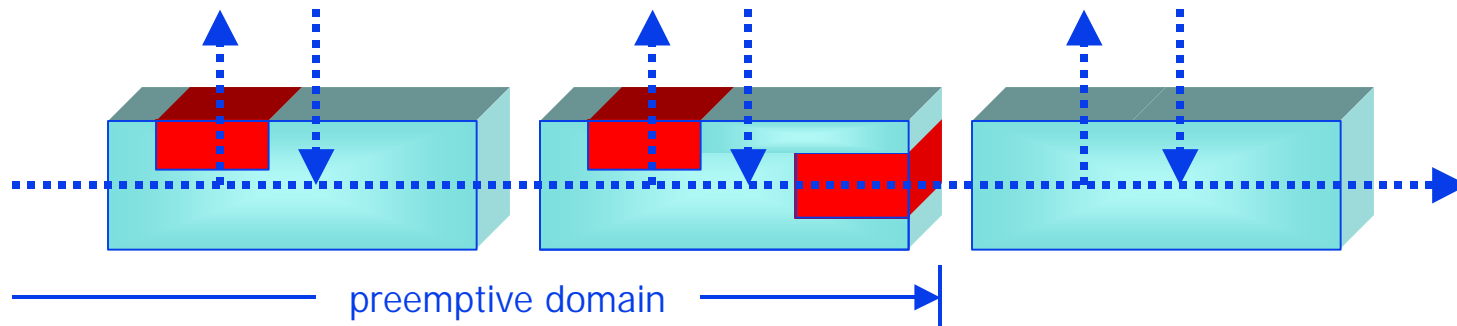
# Pre-emption fragments



- Packets can be suspended
- The class-A packet can be stripped
  - egress queues are store&forward
  - distinctive idle markers needed



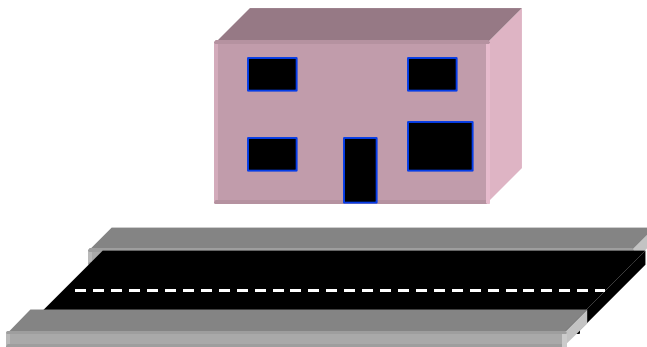
# Pre-emption compatibility



- Pre-emption mandates egress S&F
- Simplistic node has no such S&F
- Interoperability burden on elegant
  - boundary node has S&F bypass
  - cut-through in preemptive domain



# Limits of scalability



- Geosynchronous
- Terrestrial
  - The metro area
    - To the curb
      - To the home



## **Lessons of the past...**

- **Flow control mandates 2-out-of-3**
  - Low latency transmissions
  - Fair bandwidth allocation
  - High bandwidth utilization
- **Feedback control systems**
  - Low latency signaling
  - Control can pass class-B/C packets
  - Separate class-A queue is utilized
- **Other observations**
  - Local control => global perversions
  - Fairness is inherently “approximate”
  - Strange beating sequences DO OCCUR

# Allowed transmissions

|            | warnings |      | transmissions                         |     |     |
|------------|----------|------|---------------------------------------|-----|-----|
|            | LO       | HI   | none                                  | LO  | HI  |
| $\geq 3/4$ | send     | send | A,F                                   | A,F | A,F |
| $\geq 1/2$ | send     | pass | A,F                                   | A,F | A   |
| $\geq 1/4$ | pass     | --   | A,B,C <sub>b</sub> ,F                 | A,B |     |
| $\geq 0$   | --       | --   | A,B,C <sub>b</sub> ,C <sub>c</sub> ,F |     |     |



## Arbitration summary

- **Dual levels**
  - **Class-A, pre-emptive low latency**
  - **Class-B, less latency sensitive**
- **Jumbo frames**
  - **Affect asynchronous latencies**
  - **NO IMPACT on synchronous latency**
- **Cut-through vs store-and-forward**
  - **Either should be allowed**
  - **Light-load latency DOES matter**





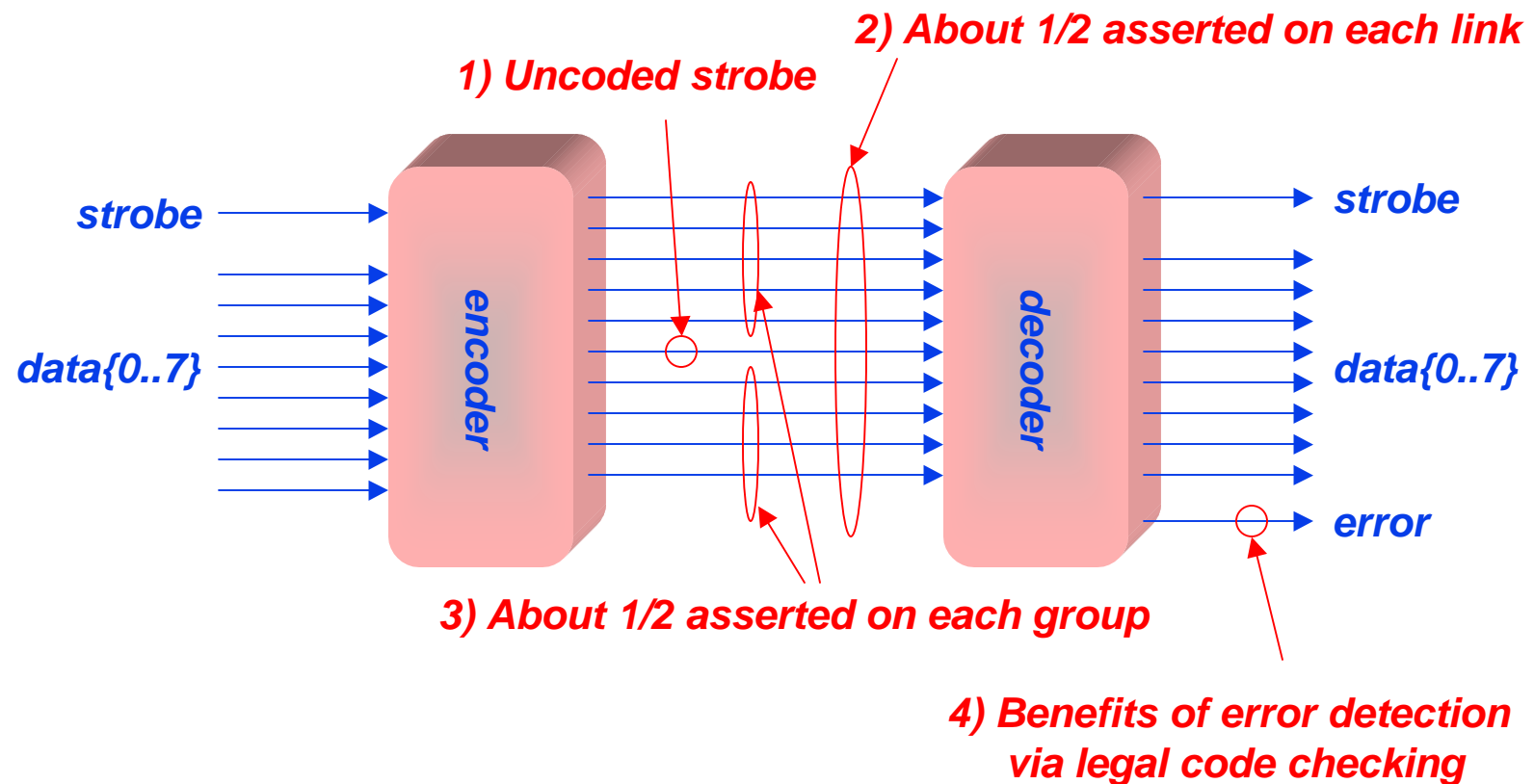
# MAC interface links

## LiteLink features

- Byte-wide 10Gb/s link (Gb/pin/s)
  - 11 pins total for clock, control, and data
- SLVS-400 transmitter, matched impedance
  - Technology independent 0.8V driver supply
  - .2V and .6V low and high signal levels
- Differential receiver, matched impedance
  - Near DC encoding (5/11 or 6/11 ones)
  - Termination derived reference voltage
- Source synchronous, DDR clock



# Byte-wide coding properties





# LiteLink signaling

