

# P1796

## Resilient Backplane Ring (RBR) slide material

### 2004Nov04b

David V. James  
dvj@alum.mit.edu

## IEEE-SA Standards Board Bylaws on Patents in Standards

### 6. Patents

IEEE standards may include the known use of essential patents and patent applications provided the IEEE receives assurance from the patent holder or applicant with respect to patents whose infringement is, or in the case of patent applications, potential future infringement the applicant asserts will be, unavoidable in a compliant implementation of either mandatory or optional portions of the standard [essential patents]. This assurance shall be provided without coercion and prior to approval of the standard (or reaffirmation when a patent or patent application becomes known after initial approval of the standard). This assurance shall be a letter that is in the form of either:

a) A general disclaimer to the effect that the patentee will not enforce any of its present or future patent(s) whose use would be required to implement either mandatory or optional portions of the proposed IEEE standard against any person or entity complying with the standard; or

b) A statement that a license for such implementation will be made available without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination.

This assurance shall apply, at a minimum, from the date of the standard's approval to the date of the standard's withdrawal and is irrevocable during that period.

## Inappropriate topics for IEEE WG meetings

**Don't discuss licensing terms or conditions**

**Don't discuss product pricing, territorial restrictions or market share**

**Don't discuss ongoing litigation or threatened litigation**

**Don't be silent if inappropriate topics are discussed... do formally object.**

If you have questions,  
contact the IEEE-SA Standards Board Patent Committee Administrator at  
patcom@ieee.org or visit <http://standards.ieee.org/board/pat/index.html>

## IEEE Code of Ethics

We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:

1. to accept responsibility in making engineering decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;
2. to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;
- 3. to be honest and realistic in stating claims or estimates based on available data;**
4. to reject bribery in all its forms;
5. to improve the understanding of technology, its appropriate application, and potential consequences;
6. to maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;
- 7. to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;**
8. to treat fairly all persons regardless of such factors as race, religion, gender, disability, age, or national origin;
- 9. to avoid injuring others, their property, reputation, or employment by false or malicious action;**
10. to assist colleagues and co-workers in their professional development and to support them in following this code of ethics.

## RBR's IEEE heritage

- IEEE Std 1212 – 1991 CSR Architecture
  - Indivisible memory-mapped update operations
- IEEE Std 1596 – 1992 Scalable Coherent Interface (SCI)
  - Busy-retry destination-based flow control
- IEEE Std 1394 – 1995 Serial Bus
  - Isochronous path reservations, time-sync, and per-cycle transmissions
- IEEE Std 802.17 – 2004 Resilient packet ring (RPR)
  - Scalable network-on-a-ring, classes of service, resiliency
- IEEE 802.3ap Backplane Ethernet Task Force
  - Physical layers for the backplane (PHYs)
- IEEE 802 (CE) Study group
  - Isosynchronous path reservations, time-sync, and frame formats

## Similar industry technologies

- Infiniband
- HyperTransport
- PCI-express
- Rapid I/O
- Others?
  - Fiber-channel, serial ATA, serial SCSI, FDDI

## Current status

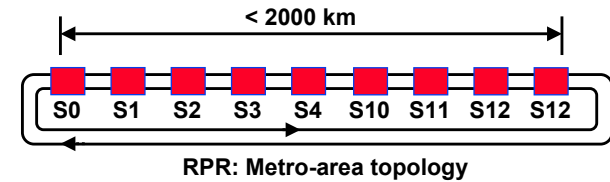
- **Study group authorized by MSC January, 2004**
- **PAR approved June 24, 2004**
- **Scope:**
  - Resilient backplane ring (RBR) is a backplane interconnect based on the dual-ring resilient topology of resilient packet ring (RPR) and the 802 MAC addressing structure. RBR includes features appropriate for the low-latency backplane environment: destination-based flow control, low-power short-haul PHY, backplane-to-backplane links, transport of IEEE-1394 isochronous data, and support of IEEE-1596 memory-update operations.
- **Purpose:**
  - The purpose of this project is to leverage the benefits of network-compatible resilient interconnects within low-latency backplane environment.

## Reasons for the RBR standard

- High speed backplanes are oftentimes used within the networking environment, where designs can be simplified by sending network frames and card-to-card communications over the same links.
- Although the resilient packet ring (RPR) has the quality of service (QoS) needed for card-to-card communications, other facilities associated with a low-latency backplane environment are missing.
- When RPR like protocols are supplemented with latency-critical backplane services, the resulting backplane interconnect should be sufficient for many mixed application backplane designs.
- Affected sectors would include enterprise networking and computer server industries; perhaps 100s or hopefully 1000s of companies.

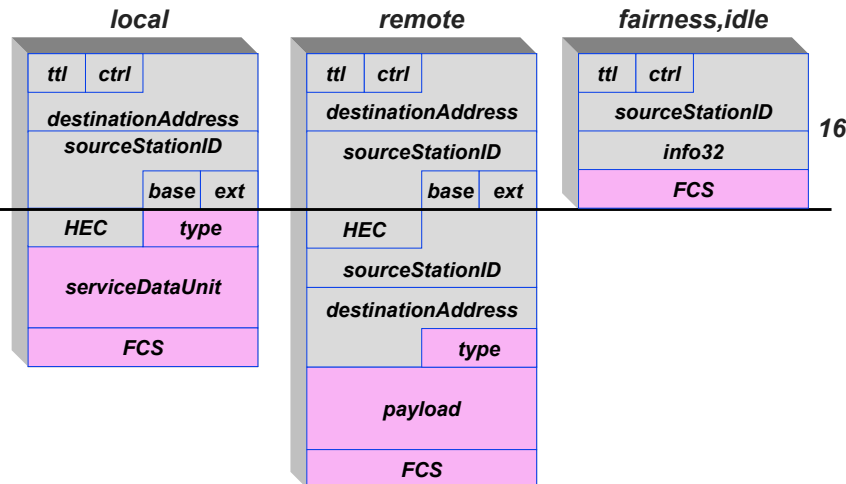
## RPR summary (a similar existing standard)

## RPR topologies



SONET environment applications  
 Duplex counter-rotating rings with spatial reuse  
 IEEE 802 frames, with ring-routing supplements  
 Several product-in-field constraints

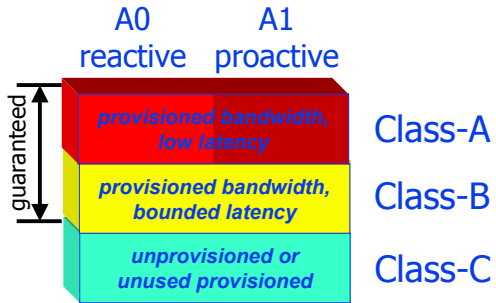
## RPR format summary



## RPR → RBR simplifications

- ~~Steering, edge wrap, and center-wrap protection.~~
- ~~Single queue and dual queue transit buffers.~~
- ~~Rate based and credit based shapers.~~
- ~~Parity, 16-bit hec, 32-bit fcs, 32-bit hec, and 32-bit fcs.~~
- ~~Conservative and aggressive fairness, without adaptive.~~
- ~~Multichoke “hooks” and single-choke specification.~~
- ~~Negotiated capabilities:~~
  - ~~wrapping mode (→ center wrap)~~
  - ~~jumbo capable (→ all are transit-capable)~~
  - ~~conservative fairness (→ all are adaptive)~~

## Arbitration classes



## RBR distinctions

- **WG development process**
  - Straw men proposals (more than slideware)
- **RBR is within smaller “boxes”**
  - Reaction times are much smaller
  - Destination-congestion: “retry” becomes practical
  - Low-powered transceivers are a must
- **RBR plug-and-play vs RPR single-vendor acquisitions**
  - Elimination of most configurable “knobs”
  - Negotiated real-time bandwidths
  - Real-time bandwidths are “real”
  - Specified time-of-day synchronization

## Residential Ethernet (RE) (an existing 802.3 study group)

## Residential Ethernet (RE)

- 1) Of course, fully supports 802.1
- 2) \*Low cost (zero additional perceived cost per unit to consumer)
- 3) MAX 500uSec DELAY THROUGH EACH HOP (illustrated by Gibson)
- 4) Performance requirements are based on 7 hops maximum (3.5 milliseconds)
- 5) \*Provide master clock with which all stations' clocks maintain bounded phase delay (illustrated by Pioneer, lip-sync and digital speaker scenarios)
- 6) \*Phase delay (between master and station clock) short term jitter maximum is within a few microseconds
- 7) \*Phase delay (between master and station clock) long term jitter (drift) asymptotically approaches zero
- 8) \*Isochronous traffic only supported over 100Mbps or greater full-duplex
- 9) \*At least 75% of link bandwidth may be reserved for isochronous traffic
- 10) \*At least 10% of link bandwidth is reserved for best-effort traffic

## Residential Ethernet (cont)

- 11) \*Assumed cycle size is 8kHz (very widely used)
- 12)\*\*Network provides a mechanism to request/assign resources for isochronous transmission (e.g. bandwidth, channel) and the default rule(s) for managing the resources
- 13) \*Isochronous packets are never dropped nor delayed due to ANY other network traffic
- 14) \*Granularity (not minimum) of bandwidth allocation is on the order of single bytes per cycle
- 15) By default, resource allocation is first-come-first-serve
- 16) \*\*Network will automatically reclaim previously allocated but currently unused resources
- 17) \*\*Isochronous traffic is not disrupted when a station/session is added/removed to/from the network
- 18) Allows bridging to isoch 802.11e
- 19) Allows bridging to isoch 1394

## Congestion management (CE) (an existing 802.3 study group)

## Congestion management (CE)

Nov, 2003: Backplane Ethernet CFI

March, 2004: Congestion Management study group  
spawned from Backplane

May, 2004: First meeting, decided not yet ready for PAR –  
still trying to understand the issues

July, 2004: First objectives

Sept, 2004: Refine objectives, PAR and 5 criteria.

Split problem into 2 areas, solve one of them in 802.1

## Congestion management (CE)

- 1) Question: When will the 802.3 CM tutorial be presented?  
Response: At: November 16, 2004 San Antonio, TX  
802 Plenary meeting 2nd tutorial slot (thought to be 8:00PM)
- 2) Question: What is the rough scoping/strategy for CM?  
Response: a) 802.3 will define a throttle that can limit the station transmission rate. This limits only the overall rate, not per-connection and/or per-class rates.  
b) 802.1 (if willing) would define the mechanisms to sense congestion and/or loss within the interconnect, to better facilitate the correct setting of the throttle rate.
- 3) Question: Is this a dynamic or static rate throttle?  
Response: It is quasi-static. Dynamic in the sense that it can vary based on (2b). Static in the sense that it controls the smoothed averaged rate. Rate updates are presumably done on an infrequent (not per-frame) basis.

## Congestion management (CE)

The 802.1 piece of the project isn't quite as you described. We're asking 802.1 to create a tag of sorts that it can use to generate FECN (Forward Explicit Congestion Notification). TCP can then use this to reduce traffic end-to-end before packets are discarded. IP already has a mechanism to carry FECN but when IP packets are encrypted, or other network layer protocols are used that don't already have a mechanism, this may be a useful addition. Some L3/L4 aware bridges already have this capability in a proprietary fashion.

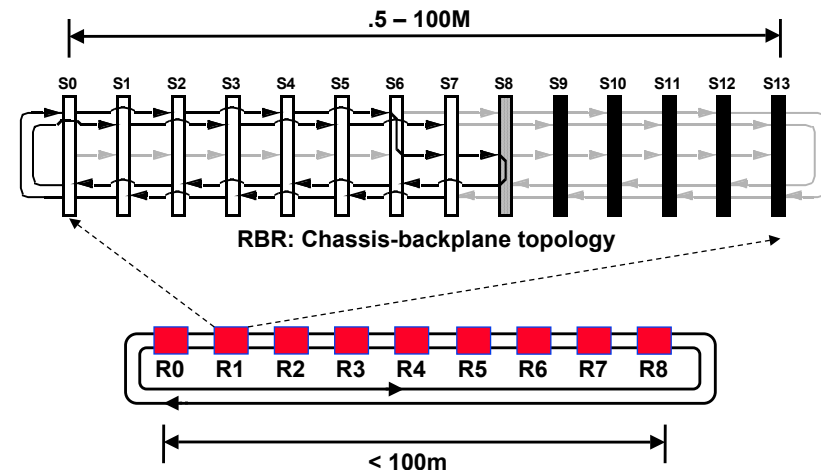
There may be other ideas that 802.1 comes up with.

## RBR summary (a starting point)

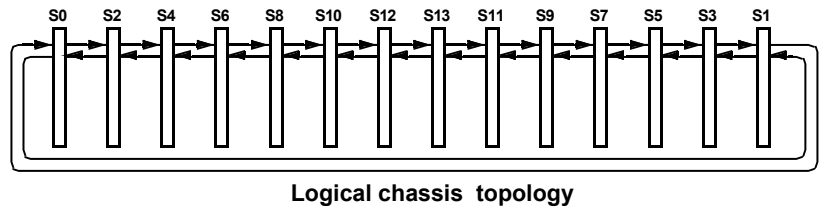
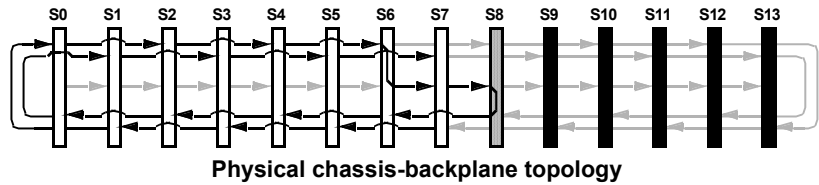
## RBR protocol summary

- Leveraged RPR values:
  - Ethernet frames with QOS delivery
  - Ring efficiency and resiliency
- QOS enhancements
  - Accurate time-of-day synchronization
  - Revised/verified classA1/classB guarantees.
  - Quasi-synchronous isochronous transfers
  - Negotiated access controls.
- Lossless transactions
  - Destination-asserted flow control
  - Hard-coded memory-access commands
  - Request/response queuing options
- Backplane PHY definitions

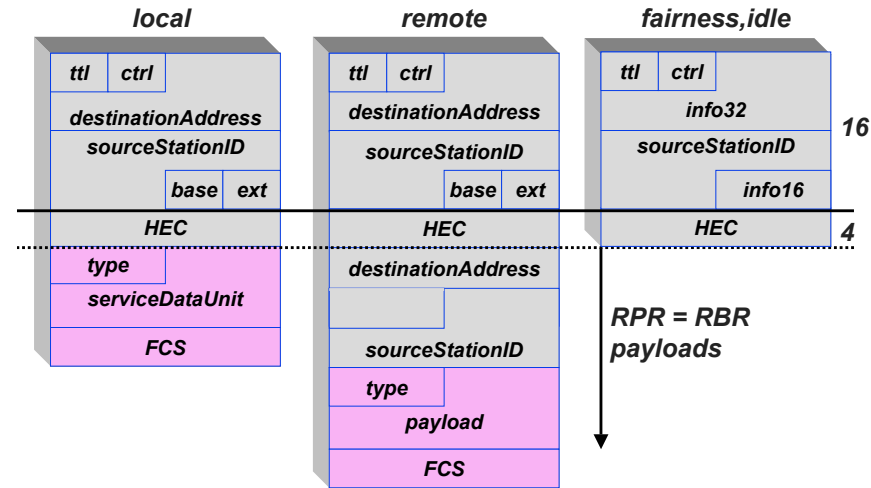
## Hierarchical topologies



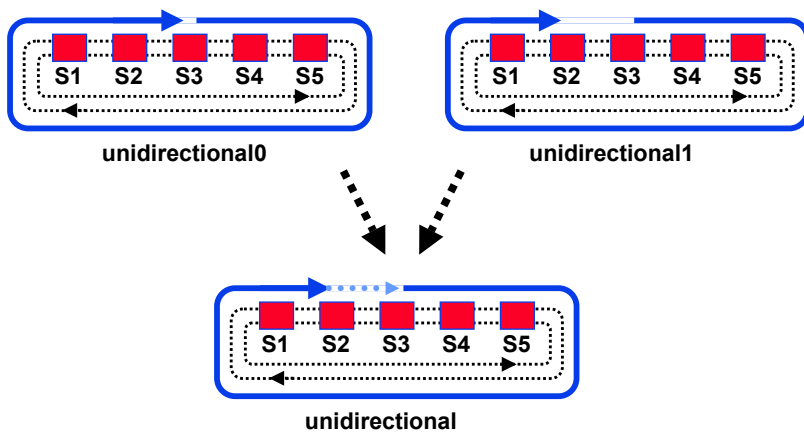
# Topology equivalents



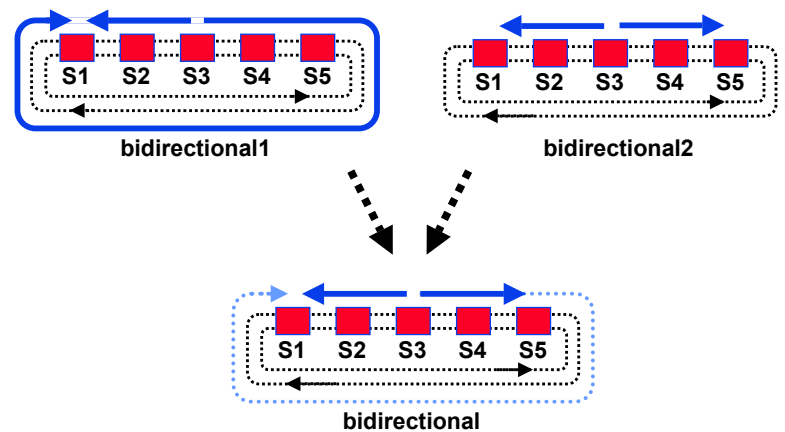
# RBR format summary



# Unidirectional flooding



# Bidirectional flooding



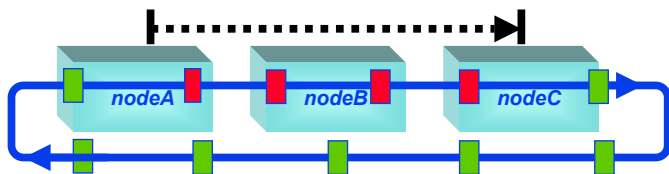
## Flow control

## Oposing arbitration



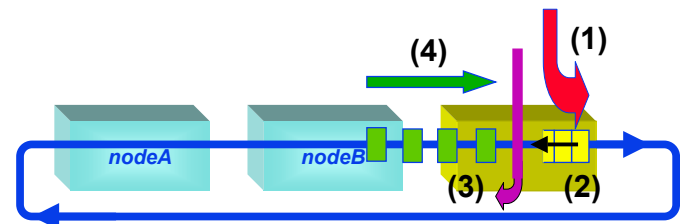
- Data packets flow in one direction
- Arbitration control flows in the other\*

## Proactive class-A0 partitions



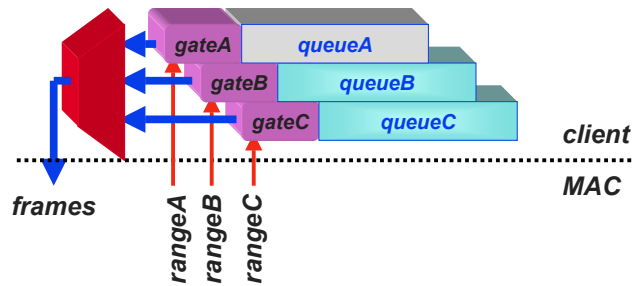
- Data packets go source-to-destination
- Residue returns destination-to-source to provide subsistence for transmissions

## Reactive class-A1 control

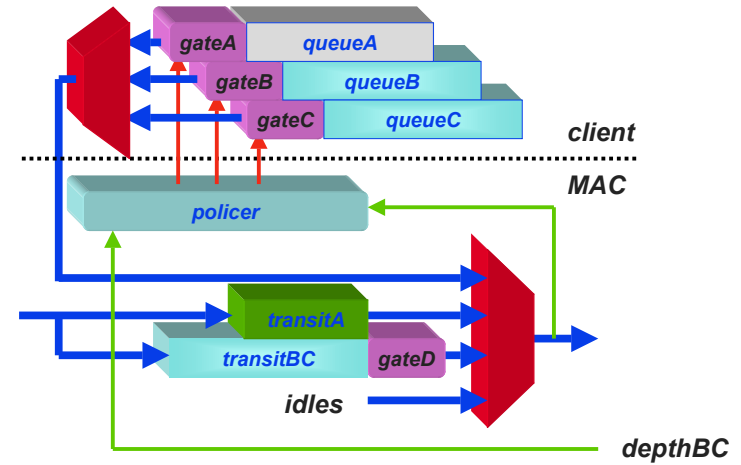


- Transmission of packets causes
- Backup of passBC FIFO that
- Returns flow-control information that
- Provides consumable idle packets

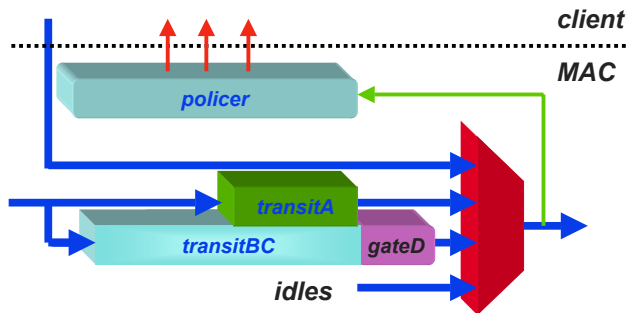
## MAC-Client interface signals



## Arbitration components

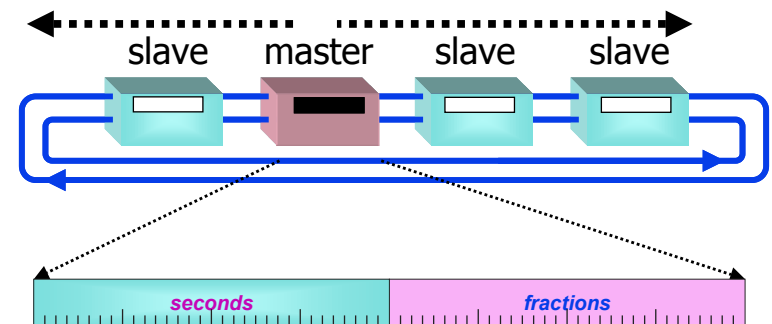


## Small-to-large transitBC



- 1) Small => proactive classA0
- 2) Medium => mixed classA0/classA1
- 3) Large => reactive classA1

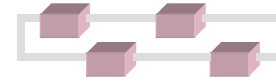
## Time-of-day synchronization (not bit-clock synchronization)



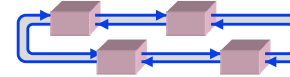
## Operations, administration, and maintenance (OAM)

- Recovery from failed components
  - “Protection” via wrap&steer
- Identify failing components
  - MIB counts
  - stomped CRCs

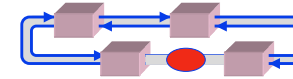
## Supported topologies



- A physical ring

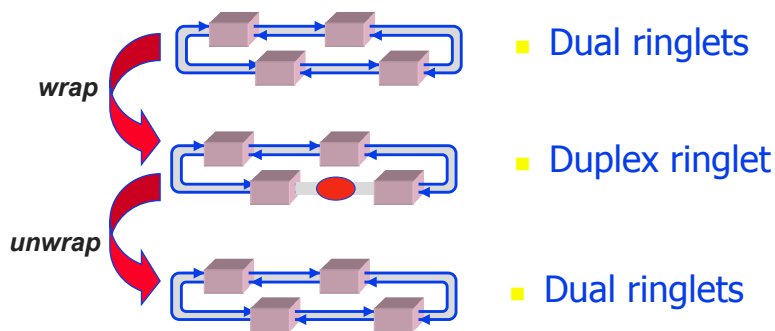


- Dual ringlets

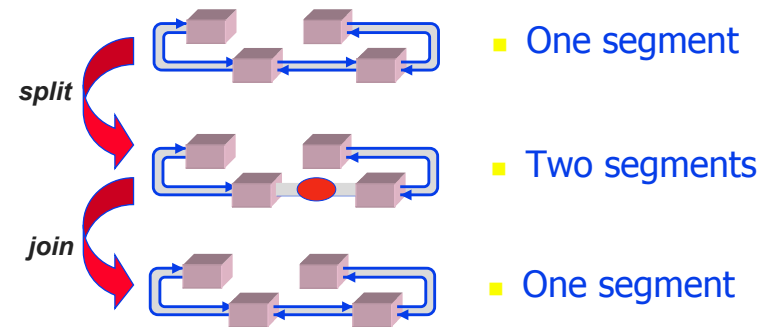


- Duplex ringlet

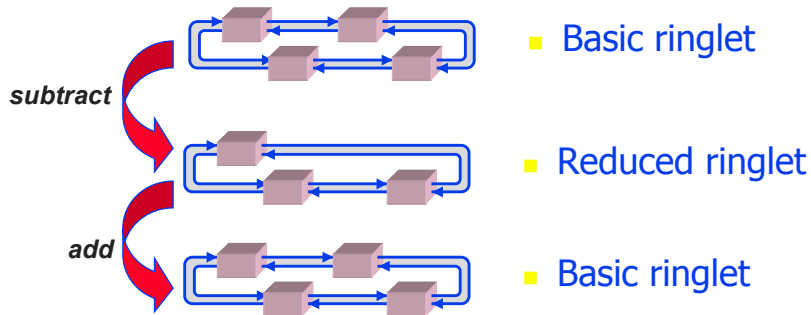
## Link failures: wrap & unwrap



## Link failures: split&join



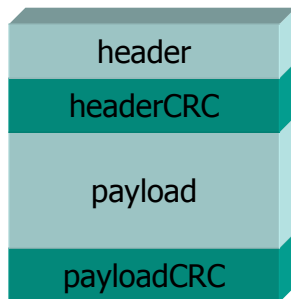
## Link failures: subtract & add



## 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

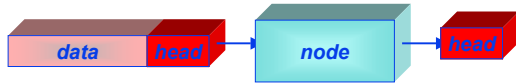


## Cut-through CRCs



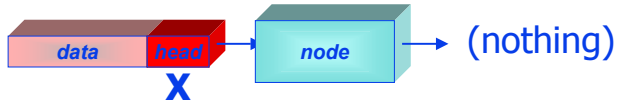
- Corrupted packet remains corrupted
- Error logged when first detected
- ```
if (crcA!=crc) {
    errorCount+= (crcA!=crc^STOMP);
    crcB= crc^STOMP;
}
```

## Distinct CRCs reduces discards



X

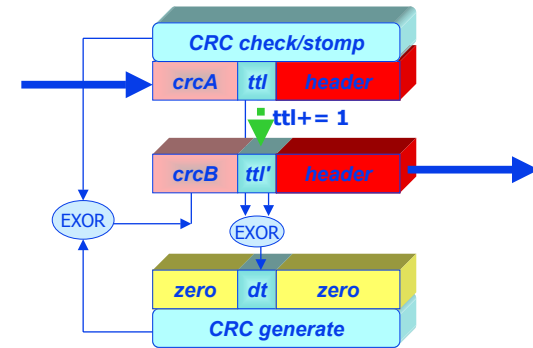
- Discard the corrupted data



X

- Discard the corrupted packet

## End-to-end CRC protected TTL



## CRC equation examples

```

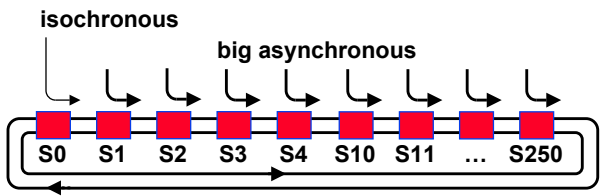
a= c00^d00;    b= c01^d01;
c= c02^d02;    d= c03^d03;
// ...
s= c14^d14;    t= c15^d15;
c00= a^        e^  g^h^    m^
c01=  b^        f^  h^j^    n^
c02=   c^        g^  j^k^    p^
c03=    d^        h^  k^m^    r^
c04=     e^        j^  m^n^    s^
c05=      f^        k^  n^p^    t^
c06= a^        e^  h^        p^r^
c07=  b^        f^  j^        r^s^

```

## Difficult remaining problems

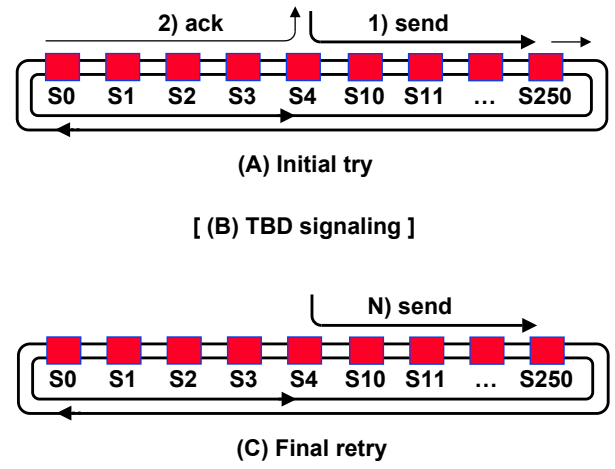
- **Classes of service**
  - Tight classA latency guarantees
  - Unconstrained classB levels
- **Destination-based flow control**
  - Busy retry has the right properties
  - per-source feedback is simple
  - output-port feedback is possible
  - Overhead must not exceed 1 retry

# Worst-case isoch delays



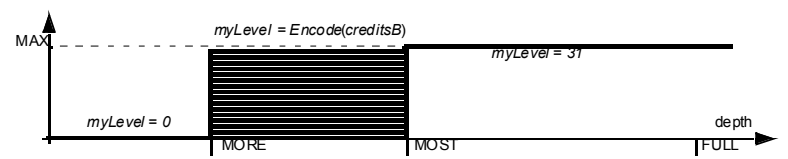
collisionDelay = stations \* MAX\_SIZE;  
 1.5kB @ 1Gb w/250 stations → 12us \* 250 → 3 ms  
 8.5kB @ 1Gb w/250 stations → 68us \* 250 → 17 ms

# Destination-based flow

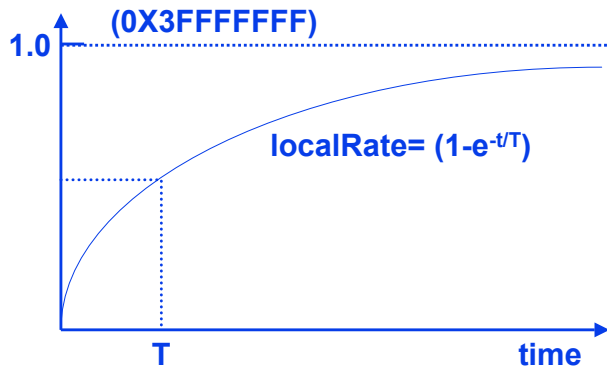


# Evolving slides

# Queue depth feedback



## DSP perspectives



## Instructions for the WG Chair

- At Each Meeting, the Working Group Chair shall:
- Show slides #1 and #2 of this presentation
- Advise the WG membership that:
  - **The IEEE's Patent Policy is consistent with the ANSI patent policy and is described in Clause 6 of the IEEE SA Standards Board Bylaws;**
  - **Early disclosure of patents which may be essential for the use of standards under development is encouraged;**
  - **Disclosures made of such patents may not be exhaustive of all patents that may be essential for the use of standards under development, and that neither the IEEE, the WG nor the WG Chairman ensure the accuracy or completeness of any disclosure or whether any disclosure is of a patent that in fact may be essential for the use of standards under development.**
- Instruct the WG Secretary to record in the minutes of the relevant WG meeting:
  - **that the foregoing advice was provided and the two slides were shown;**
  - **that an opportunity was provided for WG members to identify or disclose patents that the WG member believes may be essential for the use of that standard;**
  - **any responses that were given, specifically the patents and patent applications that were identified (if any) and by whom.**