Extending SRP Based on 802.17 Feedback

Steven Wood
Cisco Systems

Agenda

- Steering and Wrapping Co-existing
- Support for Bridging and Data Frame Formats
- Packet Error Handling
- Bandwidth Management
  - Congestion Control: Management vs Avoidance
  - Extending Fairness to Multiple Domains
- TTM for Standards Compliant Silicon

Steering and Wrapping

- Wrapping provides the fastest protection switching regardless of ring size with lowest packet loss
  - Many customers demand this feature
  - Others in .17 propose that steering is good enough
- 802.17 can support both and allow all vendors to satisfy their customers
  - Both wrapped and steered nodes can exist in the same ring
    - We will be providing a written proposal at the next meeting
    - Have to look at corner cases to make sure we are not missing something!

Support for Bridging

- Our PAR requires us to support 802.1D bridging
  - 802.17 will carry many types of traffic including:
    - IP, Ethernet, MPLS, PacketTDM, …
- Consider the Ethernet Frame as an example

DA	SA	Type/ Length	Optional .1Q	Payload	FCS

- A Proposed RPR packet format supports Ethernet bridging by simple a prepend of the RPR header
  - Note: payload could be an IP, MPLS … packet

RPR	DA	SA	Type/ Length	Optional .1Q	Payload	FCS

Support for Bridging

- No changes to Ethernet frame
  - simple mapping into .17 frame
  - RPR header check is recalculated hop by hop
    - no need to recalculate FCS at each node due to TTL
  - Inverting the FCS after error detection is a good idea
    - any errors introduced in system or ring will be caught
    - check FCS at each node and log error at first node
- Header must contain TTL, some mode, control bits and some form of Header protection
  - Is a simple parity bit enough, maybe not?
  - Anything else missing?

Support for Bridging

- Support for Unknown Unicast
  - alter packet accept logic to check if packet address is unknown and if so replicate packet into TB and copy to host
  - Topology DB can be used to fill filtering DB to determine unknown unicast
- The Source strip mechanism needs to be augmented
  - When a bridge injects a packet, the SA is stored in the filtering DB and checked to strip the packet after 1 loop.
  - TTL mechanism completes task if node goes insane
    - Requires TTL set correctly to number of nodes on ring
    - Requires TTL to be decremented by nodes in a passthru state

Why Not Protect Addresses?

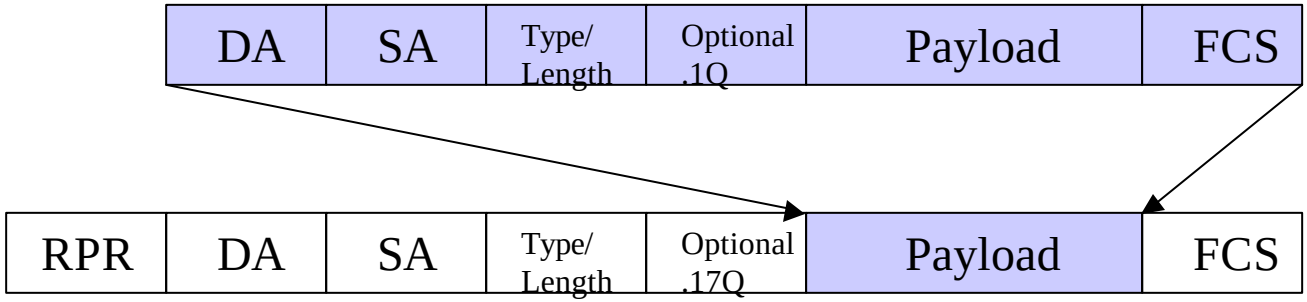
- Is it necessary to imbed the RPR header & HEC within the packet and protect the DA/SA?
  - Delivering a packet with good address and bad payload does not add a lot of value
  - Do you need to count who was going to get the packet?
    - This count is incomplete due to the cases where:
      - Error is in the header
      - Packet was dropped at a bridge due to congestion or packet error
    - Source/Destination need to count packets sent/received and then determine the loss
      - Node aggregate counting can be done for reasonable cost in the MAC, flow based counting is prohibitively expensive
- Protocols know how to deal with lost packets

Does the MAC Need Counters

- Customers have requested per node traffic measurement capability
  - Allows traffic monitoring / engineering to occur
  - Can determine how much aggregate traffic is flowing between all nodes in the ring
    - Can be used to determine exactly how many packets went missing between two nodes on the ring
  - Per flow counting is outside the scope of the MAC
    - can be implemented outside as a value add
  - Per node accounting cost in Silicon area is reasonable
    - 256 entry CAM plus counters max

Encapsulating Bridging

- Encapsulating bridging will have advantages over Transparent bridging
  - Many vendors already see this as the best solution
  - Simple mapping of the Ethernet frame into a .17 frame
  - The DA and SA of the .17 frame are from the set of MAC addresses on the ring



Encapsulating Bridging

- Advantages
  - Complete Customer Separation with optional .17Q
  - No change to Ethernet frame
    - simple mapping into .17 frame
    - RPR header check is recalculated hop by hop
      - no need to recalculate FCS at each node
      - any errors introduced in system or ring will be caught
- Disadvantage
  - Our PAR does not currently request the ability to define this, we need to finesse it

Bandwidth Management

- SRP algorithm performs congestion management rather than avoidance
  - Coupled with the SRP transit path design it provides
    - good delay and jitter performance for high priority traffic without a requirement to underutilize links or pre-provision traffic
    - Simple interface between MAC and upper layer
  - Simple implementation that allows more complex algorithms to be layered on top allowing differentiation at the box level
    - Per destination queuing not required, but can be added on top

Bandwidth Management

- SRP does not require the participation of every node on the ring in every BW allocation decision
  - Messages only flow within a local congestion domain
  - Relatively immune to lost BW allocation messages
  - Multiple non overlapping local fairness domains can occur around ring without requiring per destination queuing
  - Most avoidance schemes require the ring be underutilized by some amount
    - No such requirement in SRP

Head Of Line Blocking

- Aggregation traffic patterns do not require further optimization of per destination queues
  - What percentage of .17 traffic is aggregation ?
- Avoidance or Management schemes do not in and of themselves prevent HOL blocking
  - How to solve HOL blocking in general?
    - Requires per destination transmit queues plus
    - appropriate BW usage messages

Head Of Line Blocking

- Requiring Per Destination Queues as part of the MAC (or above it) is not reasonable and should not be part of the standard
  - Forces everyone to add additional HW (cost)
  - Removes differentiation and commoditizes the box
- Allow people to choose whether to add this complexity as part of their differentiation

Extensions to SRP

- The SRP algorithm (today) does not support the concept of per Destination Queues
- It can be extended to do so
  - BW allocation message carries the source address of the node that is experiencing congestion (choke point)
    - Today this message only travels partway around the ring
    - Allow it to travel the entire ring, so all nodes know about the choke point and how much BW can get through it
    - Queuing chip above the MAC then controls ingress into the ring based on where the packets are going and the choke point information
    - MAC does not rate limit above traffic based on old FA

Extensions to SRP Fairness

- Other extensions are possible
  - If they provide improvements to real problems they should be explored
  - We would like to work in 802.17 to investigate:
    - Minimizing the size of the transit buffers
    - Layering of per Destination Queues on top of basic BW management
    - Shaping of ingress traffic as part of the MAC to provide better jitter/delay
    - Algorithms and HW extensions that could allow both styles of transit path design to co-exist
      - Transit path implementation specifics is not a part of the standard
      - Single transit queue and multiple transit queues can co-exist
  - Some of these changes will not be backwards compatible to current SRP implementations

TTM for Standard Si

- SRP is currently the only fully described proposal
  - Definitely the most studied and deployed
  - We need this level of description and study of all proposals
- Changes to SRP can insure a level playing field for everyone
  - 802.17 Si can be available next year by multiple vendors
  - Everyone gets access at the same time