A Portland Proposal[1]

This proposal is an attempt to address diverse interests in the 802.17 working group. It is unlikely that anyone will find it acceptable in its entirety, but it would be interested to understanding how people would change this proposal to suit their requirements. This document is posted in Word and .pdf so people can respond with a modification of the document if they choose. Comments can be sent to bob.sultan@fnc.fujitsu.com or posted to the working group reflector (stds-802-17@ieee.org). If you send your email to the reflector, please don't attach the file.

It is intended that the terms used in this proposal correspond to those defined[2] in the *terms and definitions* document (for latest version see http://www.ieee802.org/rprsg/public/presentations/jul2001/index.html). Questions about terms can be sent to the terms and definitions reflector (stds-802-rprsg@ieee.org). The terms have been italicized in this document.

- *dual-ring*[3]
- standardized *MAC protocol interface* between neighboring *stations* on the *ring*[4]
- *PHY agnostic*[5]
  - *reconciliation sublayers* for *SONET* (*OC-3* and above) and *Ethernet PHY* (1G and above)
- *protocol agnostic*[6]
  - insensitive to the *protocol type* of the *MAC client*
- *payload agnostic*[7]
  - insensitive to the *payload* supplied by the *MAC client*
- *unicast* to *MAC individual address* with *frames stripped* at *destination station*
- *multicast/broadcast* to *MAC group address* with *frames stripped* at *source* after circulation
  - *multicast registration* similar (or analogous) to that provided by *GMRP*
- three[8] *traffic classes*[9]
  - ***synchronous traffic class (STC)***
    - fixed *frame size*[10] (e.g. 64 bytes)
    - *buffer insertion ring*[11] *(BIR) for transit*

---

[1] Version 6/22/02. This document is intended to describe one of many possible RPR architectures. It is not intended to describe RPR objectives or requirements.
[2] Some are missing, but will be added in a future version.
[3] The method of operation of a multi-ring (ie. multiple concentric ringlets) is not clear. If there is a requirement for a multi-ring, a specific proposal should be described. Link aggregation in the style of 802.3ae can be deployed, but the scope is between stations on the ring to the duality of the ring is not affected.
[4] Implies that stations are fully interoperable with respect to all activities associated with shared use and access to the ring medium.
[5] As is any MAC layer conforming to the OSI model (we favor removing these statements regarding agnosticism)
[6] As is any MAC layer conforming to the OSI model (we favor removing these statements regarding agnosticism)
[7] As is any MAC layer conforming to the OSI model (we favor removing these statements regarding agnosticism)
[8] Variations on these classes might increase the number beyond three. For example, there could be distinct synchronous traffic classes for fixed size frames of 64bytes and for 256bytes.
[9] The traffic classes are visible at the ingress and at transit stations (are they service classes at the ingress?). Disclaimer: the descriptions of the *traffic classes* are likely to have bugs, particularly the *synchronous traffic class* that we added to meet requirements that we have heard expressed by others.
[10] Could have multiple synchronous traffic classes, each with a different fixed size frame (e.g. 64, 128, 256).
[11] Implies fixed buffer size that is small multiple (e.g. 6, 3, or 1) of fixed frame size (see terms and definitions).

        - worst-case per-*station transit delay* of one *frame_time*
        - worst-case *end-to-end delay* of *(#stations X frame_time)*
    - *access queue* holds *(previously shaped) ingress*[12] *traffic* waiting for gap in *transit traffic*
        - *transmit* from *access queue* only when *insertion buffer* is empty
        - worst-case *access delay* is *#stations X MTU_time*[13]
        - *access buffer size* is *#stations X MTU_size*[14]
    - *ring end-to-end delay* is bounded
        - since *access delay* and *transit delay* are bounded
    - intuition: traffic of this class is assumed to be *shaped* prior to arrival from the *MAC client*. *Shaping* reduces the worst-case number of contiguous *frames* in *transit* and thereby reduces worst-case *access delay*
- ***guaranteed traffic class (GTC)***
    - *buffer insertion ring*[15] *(BIR)* for *transit traffic*
        - worst-case per-*station transit delay* of one *MTU_time*
        - worst-case *end-to-end delay* of *(#stations X MTU_time)*
    - *shape ingress traffic received* from *MAC client*
        - *shaping buffer* of size *MBS*[16]
            - relatively uniform spacing between *frames inserted* by a *station*
            - guarantees number of contiguous *frames* on *ring* is no more than the number of *stations*[17]
            - allows *access delay to be bounded (*see *access queue* below*)*
    - *access queue* holds *ingress traffic* waiting for gap in *transit traffic*
        - *transmit* from *access queue* only when *insertion buffer* is empty
        - worst-case *access delay* is *#stations X MTU_time*[18]
        - *access buffer size* is *#stations X MTU_size*[19]
    - *ingress buffer* is the combination of *shaping buffer* and *access buffer*
        - *end-to-end delay* is bounded
            - since *ingress shaping delay*, *access delay*, and *transit delay* are all bounded
    - intuition: shaping on ingress reduces worst-case number of contiguous transit frames and thereby reduces worst-case *access delay*
        - some *frames* experience more *delay* than they need to, but it's worst-case *delay* that's important for *SLA*
- ***best effort traffic class (BETC)***
    - no guarantees
    - *ingress traffic* shaped in accordance with the *allowed ingress rate*
    - *weighted fairness*
        - *simple fairness* does **not** make most effective use of uncommitted *capacity*
            - effective use of *available bandwidth* is objective of *BETC*
    - *fairness algorithm* described in RFC 2892 can be used, assuming:
        - enhancement for *weighted fairness*

---

[12] 'Ingress' is the direction from the MAC client towards the ring (see terms and definitions). There is distinct ingress queuing per ringlet.
[13] If we assume that the shaping of guaranteed traffic on ingress is not coordinated with the already shaped synchronous traffic, then the worst-case delay is 2 X *#stations X MTU_time.*
[14] If we assume that the shaping of guaranteed traffic on ingress is not coordinated with the already shaped synchronous traffic, then the buffer required is 2 X *#stations X MTU_size.*
[15] Implies fixed buffer size that is small multiple (e.g. 6, 3, or 1) of MTU size (see terms and definitions).
[16] A smaller buffer can be used if loss at ingress is tolerated.
[17] Assuming <100% ring bandwidth committed
[18] If we assume that the shaping of guaranteed traffic on ingress is not coordinated with the already shaped synchronous traffic, then the worst-case delay is 2 X *#stations X MTU_time.*
[19] If we assume that the shaping of guaranteed traffic on ingress is not coordinated with the already shaped synchronous traffic, then the buffer required is 2 X *#stations X MTU_size.*

    - no problem discovered in simulation
- Once in progress, the *transmission* of a *frame* by a *station* proceeds to completion.
- No specific upper limit on *MTU size,* but compliant *stations* allowed to drop frames > 15xx bytes
  - allows deployment of *rings* such that all *stations* support large *MTU*
    - greater *end-to-end delay* and requirement for larger *buffers* implied by such deployment
- CRC check performed at transit and destination stations (e.g. to allow adjustment TTL increment)[20]
- CRC computation performed at source and transit stations[21]
    - does not prohibit deployments where a *station* does not check the *CRC* (allowing *cut-thru*)[22]
- Conforms to *IEEE 802.1D transparent bridging*[23].
  - does not preclude interconnection at *L3* or other types of L2 interconnection.
- **GFP**[24] *core header* with *payload type* unique to *RPR* (ie. **not** '*ring*' type)
  - avoid '7E' expansion
  - consistent with carrier requirements for *GFP*
  - *payload* (*802.17 frame*) generally similar to *802.3 frame*
    - with *RPR* specific information prepended
      - 48-bit *802.3 MAC addresses*
      - 16-bit (or 8-bit, or 4-bit) *port number* identifying a specific *MSAP* within the *station*[25]
    - optional *VMedia tag* similar to *VLAN tag*
    - TTL
- 50 ms. *ring restoration time* in the event of a single *station* or *link* failure
  - *wrapping* is described by the standard but is a *station*-specific option
  - all *stations* must support *steering* and coexist on the *ring* with *stations* that additionally support *wrapping*[26]
- *optional VMedia* analogous to the *802.1D VLAN*
  - partitions the *ring medium* into multiple *virtual ring media*
  - optionally associates *MSAP* with a *VMedium* or multiple *VMedia*
    - binding established dynamically (*GVMRP* analogous to *GVRP*)
  - limits scope of *broadcast traffic* to *VMedium*
  - supports separation of customer traffic on ring
- supports *plug-and-play* operation
  - *ring* returns to normal operation after *insertion/removal* of *station* without *provisioning*
  - inserted *station* can *source/sink best effort traffic* without *provisioning*
    - for *station MAC,* no *VMedia*, and equal *weighting*
- defines *QoS parameters*
  - analogous, for example, to that specified for *frame relay* by *ITU I.233.1*
  - applied to a *flow* (vs. *connection* as in case of *frame relay*)
    - *flow* describes *traffic aggregate* from one *MAC* to one or more specified *MACs* on the *ring*
- provides means to specify *MAC addresses* and *QoS parameter values* associated with a *flow*
- specifies *service level definitions*
  - analogous, for example, to those specified for *frame relay* in *FRF.13*

---

[20] We are open on question of CRC at transit nodes.
[21] We are open on question of CRC at transit nodes.
[22] As in 802.3/802.1D, cut-thru is not addressed by the 802.17 standard. Cut-thru can reduce average delay, but it does not reduce worst-case delay, as cut-thru for all frames cannot be assured.
[23] Required by the 802 committee.
[24] There may be some benefit in the GFP-style frame delineation. Otherwise, details of the frame format are not likely to have a big impact on RPR function or performance (an opinion).
[25] Used to distinguish different MAC clients (e.g. subscribers) associated with the station. It is also possible to identify each MSAP with, for example, a 64-bit MAC address, the first 48-bits of which resembles a 48-bit MAC address and the remaining 16-bits identifying the specific MSAP (port) within the station.
[26] Cisco has offered to provide a description of how this coexistence works.

- specifies *operations, administration, and maintenance (OAM)*
  - analogous, for example, to that specified for *frame relay* in *FRF.19*
- specification consistent with other *IEEE 802 standards* unless specific reason for divergence
- *layers*, *protocol interfaces*, and other abstractions conform to *ISO X.200*
- *MAC station* structured to allow deployments that do not support all *traffic classes* at *ingress* but coexist with nodes that do
  - e.g. *802.17* interface card for router interconnection that supports *ingress* for *BETC* only
    - supports *transit* and *control* for **all** *traffic classes*
    - does not preclude deployment of a *ring* with *stations* supporting only *BETC*
- *station* structure sufficiently modular to allow alternative *MAC subsets* in hardware devices

Appendix: items explicitly not required

- *Maximum ring latency* specification
  - the *ring latency* requirement is established during the network design phase, considering
    - maximum allowed *end-to-end delay* for application
    - other contributions to *end-to-end delay* (propagation, packetization, etc.)
  - provider may choose **high** *ring latency*
    - for lower equipment cost (e.g. cheaper switch or memory)
    - for larger number of *stations* on the *ring*
    - for longer distances
    - for lower *data-rate rings*
- no requirement for *circuit emulation traffic class* (ie. lower *delay* than *synchronous class*, requiring *preemption*)
  - *synchronous traffic class* bounds worst-case *delay* and is appropriate for real-time applications (e.g. packet voice) except in case of
    - low *data-rate* (e.g. < 1 Gbps.) and/or many *stations*
      - but current low *data-rate rings* generally have few *stations*[27] (e.g. 3 to 6)
      - migration to data-rates above 1Gbps in *MAN* makes this an interim case
    - long distance (*WAN*)
      - large *propagation delay* is dominant component of *ring end-to-end delay*
      - difference in *delay* between *synchronous traffic class* and *circuit emulation traffic class relatively* small
  - Ring latency appropriate for packetized voice can be achieved using the synchronous or guaranteed traffic classes
    - assuming that low *data-rate rings* have few *stations* (our experience)
    - assuming that the *ring circumference* is not so large that *propagation* delay is very great relative to the *end-to-end delay* budget
- no minimum frame size[28].

---

[27] Also, < 1 Gbps. MAN likely to be SONET (vs. Ethernet PHY) in which case TDM for real-time traffic on a separate STS is a solution for low data-rate legacy deployments.

[28] Why is this needed?